Simulating the Spread of Pertussis in Enschede Region Using Agent-Based Modelling

Shaheen Abdullah Abdulkareem
March, 2010

Simulating the Spread of Pertussis in Enschede Region Using Agent-Based Modelling

by

Shaheen Abdullah Abdulkareem

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: (fill in the name of the specialisation)

Thesis Assessment Board

Chairman:	Dr. O. Huisman
External examiner:	Dr.Ir. D.F. Ettema
Supervisor:	Ms Ir. P.W.M. Augustijn
Second supervisor:	Dr. J. Flacke

ITC

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

Abstract

Epidemic disease out breaks is considered one of the main threats that have its impact on development in all countries as it has no geographical boundaries .There is a need to represent it geographically and analyze the spread of diseases using modelling tools.

Simulating the human's interactions is a complex process in general. Furthermore, the results of these interactions, which cause the spread infectious diseases is complex and required a lot of details and accuracy in simulating such phenomenon. Agent based modelling takes into consideration the spatial and temporal aspects of disease diffusion. Besides, it models the individuals contact process by contacting infected person or carriers and behaviour that affects this diffusion.

The aim of this study is to create a prototype of geographically explicit agent based disease modelling for the spread of pertussis in the Netherlands covering some aspects of this complex social phenomenon; it provides a social interaction model integrated with a disease model to simulate the spread of pertussis in Enschede city.

Agents movement over space from location to another based on activity time table according to their age. The pertussis disease model is built based on an existed model called Individual Space-Time Activity-based Model (Yang and Atkinson 2007; Yang, Atkinson *et al.* 2008), and the principle of its disease model based on the simulation of pertussis presented by Hethcote (1999a). The city environment is an explicit GIS space containing the locations where agents are going to apply their activity patterns and interact with the other agents. Environment and agents has no direct interaction or effect on each other. The City of Enschede was selected as a study area. This city is a suitable study area for a pertussis model as it is large enough to meet the CCS (critical community size). The study built synthetic population of Enschede city, and provided its agents with same basic daily life activities to participate in. The result of their participation in their activities led the disease spread. The geographic spread of the model output (diffusion of the disease) was depending on the participation of infectious individuals in activities and interact with susceptible. Yet, within given results of this model, it is still important to simulate this disease in a geographically explicit way because of individuals are spreading the disease while they are moving from place to another. Currently the map extent of the simulation only includes part of the Enschede area. The research showed a prototype of a geographically explicit simulation for the Enschede region that is able to simulate 60,000 agents, at a time interval of 30 minutes for a longer period of time (month) using Repast Simphoy. The current spread of the disease shows a reproduction rate that is comparable to values found in literature.

Keywords: Agent-based Modelling, Activity-based, infectious disease spread, pertussis, synthetic population, disease simulation.

Acknowledgements

“ *Read: In the name of thy Lord Who createth, (1) Createth man from a clot. (2) Read: And thy Lord is the Most Bounteous, (3) Who teacheth by the pen, (4) Teacheth man that which he knew not. (5)* ” Al-Alaq

I thought it would be easy to write acknowledgment but now I am trying to find word can express my grateful to everyone help me to let this thesis finish. The first one I should start with is Ms. Ellen-Wien Augustijn my first supervisor who it is honour for me to call her my mother-in-thesis for her kind, her patience, strong and friendly support for me that are real gifts from God to me I would never forget. I am proud to be in your team of research Ms. Ellen-Wien. I would like to thank my second supervisor Dr. Johannes Flacke for his support and suggestions. Great Thanks also to Mr. Bas Retsios the ITC RepastS expert and best programmer ever seen for his unbounded help to me spatially in my implementation. I should say thank you Dr. Otto Huisman for joining the team and providing me with advices and books. Dr. Ivana Ivanova, thank you for your nice smile and support since proposal time.
I should thanks the librarian staff for their kind helps in providing me articles and material for reading. Great thanks to all my GFM 2009-2010 classmates whom I spend a very nice time with them. Proudly I can say I have brothers and sisters all over the world. I should say thanks Panisara Daorueang for providing me the data of Enschede.

I have to say thank you Jowan and Rasha for their taking care of me, shopping and cooking for me. Really thank you girls

To study in Europe was not be true for me unless Erasmus Mundus scholarship program was not help me by giving me this chance to be one of the ITC MSc students. Thank you EM lot 7. It is a nice occasion to thank Ms. Angelique Holtkamp for her encouraging me to come and studying in Netherlands.

Last not the least but indeed they are the reason of my life, my mother and father who supported me along my life and nowadays with their love, prays, wariness and kindness. What to say to you both? Nothing can express what you did and do for me... I love you and I am proud of you.

“*unless Allah willed. We raise by grades (of mercy) whom We will, and over every lord of knowledge there is one more knowing.* “

Table of contents

List of figures

List of tables

1. Introduction

1.1. Motivation and problem statement

One of the important issues in any society is health because of its direct impact on humans. Keeping the population healthy is a crucial part of the mission for any government around the world. However, everyday there are break headlines in newscasts and newspapers talking about an epidemic somewhere on the globe. People follow such news with great curiosity. They track the diffusion and movement of these outbreaks when they explode and become a global event. For example, Mexican flue, which started in Mexico rapidly developed into a global outbreak that most, if not all, people from all age groups around the world are at the risk of being infected with.

Epidemics have no geographical boundaries and do not distinguish one society from another. They are considered a main problem in developing countries as well as other countries. Preventable infectious diseases are the controlling threat in the developing world. Societies there are suffering from the significant mortality rates as well as the adverse impact of illness on their way toward development. However, “the most active areas of pandemic influenza transmission currently are in parts of central, eastern and south eastern Europe, North Africa, and South Asia” depending on updating reports of World Health Organization WHO (WHO-Website 2010).

Generally, outbreaks or epidemics are often unpredictable and inevitable events. For analyzing the spread of diseases, simulations are very useful. Epidemic prediction depends on the quality of the diffusion model of epidemics in spatial population and the motion of individuals in different levels (Colizza, Barthélemy *et al.* 2007). Simulation can result in a better understanding of the transition mechanisms as well as provide practical advantages like the capability to establish and experiment theories, design and analyze epidemiological surveys, predicate trends and experiment control measures (Yong Yang and Atkinson 2008). Based on measuring the scale of current epidemic threat such as influenza, HIV etc., and on planning an effective fast policy response, it is important to have epidemic models(Epstein, Cummings *et al.* 2004).

Disease spreading is assumed to be a social phenomenon that requires tools and models to simulate such phenomena. Diseases simulations describe the transfer of the disease among susceptible and infected individuals. Disease simulations consist of an infection model and a social interaction model. For a long time , scientists presented many epidemiological models to simulate disease diffusion such as the classic mathematical equation approaches and complex systems approaches.

One of the best choices to simulate such phenomena is the use of agent-based modelling (ABM). ABMs provide effective tools to model the spatial-temporal aspects of the diffusion of diseases, model the individuals contact process at fine scale, and model the individuals’ behaviour effects on this diffusion.

One of the key factors that determine the risk of infection is the social interaction between individuals (not the distance between them). Interactions are required for an infectious disease to diffuse; here the interactions are the contacts with infected person or carriers. Different diseases seek different levels of interaction between people to effect transition (Read, Eames *et al.* 2008).

The large number of interactions an individual has constructs a network called social network. This social network consists of individuals and relations among them shaped in what are called nodes and links. The social network approach handles the structure of relationships and the interactions between individuals (Wey, Blumstein *et al.* 2008). Tracking interactions between individuals in relevant social networks and geographic areas guide to feedback patterns and phenomena that can be studied such as the diffusion of news, knowledge and infectious diseases (Epstein, Cummings *et al.* 2004; Cointet and Roth 2007).

Unfortunately, within the existing ABM disease simulations, there are a couple of problems in relation to "social models”. Some of the current models do not include an explicit spatial representations such as the country-level model (Epstein, Cummings *et al.* 2004). On the other hand, they contained an explicit spatial representation but it is not a geographic spatial representation such as the virtual city models EpiSims and the later versions of the model (Barrett, Eubakn *et al.* 2005; Eubakn 2005; Toroczkai and Guclu 2007). ABMs might exist with an explicit geographical space but with no interaction model such as the ABM model presented by Perez and Dragicevic (2009) or they have been provided with social interaction models but for non applicable disease such as HIV or the models of BioWar attack (Carley, Altman *et al.* 2004)

1.2. Research Identification

1.2.1. Research objectives

The main objective of this research is to create a prototype of a geographically explicit agent-based disease model for the spread of pertussis in the Netherlands including a social interaction model.
In order to get the main objective the following sub objectives were addressed in the research:

- Creating a synthetic population.
- Study and select an existing agent-based disease model and implement the interaction model.
- Create a disease model.
- Create a social interaction model.
- Test the model using empirical Pertussis data.
- Verification and validation the model.

These sub objectives divided to couple of questions. These questions will draw the pattern of achieving these main and sub objectives.

1.2.2. Research questions

The research questions consists the following:

1. How the synthetic population will be created?
2. What data and methods will be used?
3. Which existing agent based simulation will be used?
4. How do these models simulate the social interactions and their aspects?

5. How the disease model will be designed and created?
6. How will a social interaction model designed?
7. How the designed model would be tested and evaluated?
8. What dataset will be used for testing and evaluating the model?
9. How will the result of the simulation compare with empirical data?
10. How will the simulation output verified and validate?
11. What will the final model depict and visualize as a result of running the simulation?

The above questions will be answered during the chapters of the thesis. The following section provides an overview of the method been used.

1.3. Research Methodology

In order to achieve the research objectives and answer the questions the research will follow a methodology. The methodology adopted for the research includes:

- Literature study includes the study of ISTAM and pertussis simulation models:
    - Understand the concepts used within ISTAM. This includes synthetic population, activity based models and grouping concepts
    - Learn and understand the pertussis vaccination simulation presented by Hethcote (1999a) which is applicable for the Netherlands.
    - Connect the both models to produce a model simulating the spread of pertussis.

- Acquiring skills in Repast Simphony.
- Create Conceptual design, which is including:
    - construction of the synthetic population
    - the social interaction
    - disease spread
- Data preparation
    - Looking through the literature and other sources to extract the data that is needed for the simulation.
    - Analyse the vaccination data for Netherlands.
    - Analyze the empirical data of pertussis cases recorded during the pertussis outbreaks.
    - Looking for the common activity Dutch people used to performed during their normal daily life.

- Design implementation using Repast Simphony (see the stages below)
- Verification the model

The stages of the implementation include:

First stage: creation of Enschede synthetic population

In this stage, the model builds the synthetic population for the model using the census data of Enschede. Create the population agent-by-agent per neighbourhood starts from the household head then the partner if exist. They have been given their attributes values in addition their family id and address. After creating the neighbourhood children, they are distributed over the created households.

Second stage: load the city Environment

In this stage, the geographical environment space will be loaded to the model. This geographical space consists of GIS layers. Two types of layers exist within the model: layers that are part of the simulation calculation and layers that are uploaded for the visualization purposes. The first type of layers consists of buildings, which are the activity location where the agents aimed to perform her/his activity there.

Third stage: load agents to the environment

In this stage, agents are loaded to the environment. They have already been providing with their home addresses. Then they should look for their activity locations. Saving the location coordinates in their memory for the next visit.

Forth stage: establish agents' relationships

In this stage, the agents will look for friends and relatives. They have been provided with roles of choosing and selecting individuals to establish their social relationships. These social relationships are a part of their social interaction and activity pattern. These friends will be used for performing the activity of visiting friends or performing a group activity.

Fifth stage: Enable agents in their environment

A: Daily activity pattern

When the simulation is starting, individuals start their activities and applying their behaviours of joining groups and interacting with the other individuals. They start their day from their homes then change their position and group within the flow of the time. Within the group, they interact fully or partially with the other group members. At the end of the day, they back again to home and staying there until the next day. Along one-month individuals applying their activities changing between weekdays and weekends activity, copying the activity of a week on next weeks.

B: Disease Spread

During the day and performing activities, when agents join a group they might meet with sick individuals in the stage of infectious. Based on their immunity level they may become infected with the disease or not. After being infected, they continue their normal activity pattern for seven days before they have been forced to stop their normal activity and starting their illness activity, which is staying at home until recovery day. Within these seven days coming after infection, the first three days the individual does not spread the disease to surrounded individuals. By the starting of day four they become infectious and infecting individuals sharing the space-time with him/her. This infectious period continues until day number seven when they stopped moving from place to another. After the stopping the will not spread the disease to any individual in the simulation.

1.4. Organization of the thesis

The organization of this research thesis comes with the following sequence after this chapter of introduction:

- ⇒ Chapter 2 is addressing the theories and background of the research. In addition, it provides the reader with brief details about some concepts being used during the research. Furthermore, the chapter tries to provide some theoretical answers for the research questions. An overview about the related works and previous works can be founded within this chapter.
- ⇒ Chapter 3 is presenting the model conceptual design with the related aspects and assumptions made to build pertussis model. It has been provided with some diagrams explaining the architecture of the model to give the reader better understanding about the model.
- ⇒ Chapter 4 is discussing the required data for the model. Data captured from what type of sources and the preparation needed for these data also provided within the chapter. Moreover, the processes of preparing these datasets are presented in chapter 4.
- ⇒ Chapter 5 presents the implementation of the conceptual design. It starts with an overview about the tools and the concepts that are used for the implementation. Then the stages of coding the conceptual design are presented one by one.
- ⇒ Chapter 6 presents and discuss the results of implementing and running the conceptual design then analyzing the results using ArcGIS tools and time cube to present the results. Verification, validation and calibration of the model can be found with this chapter too.
- ⇒ Finally, chapter 7 comes with the conclusion of whole the research. The conclusion is constructed from the results of each step of the model and the influence of these steps is provided too. The chapter addresses some recommendation for the next coming work within this model. \

2. Background and Theories

2.1. Modeling the Real World phenomenon

Representing things in the world is called modelling. Modelling allows ideas to be developed and investigated. It can give a meaning of understanding the problems concerned in building something. It is the communication lines the developer and user. Models provide a component of methods that are used in developing activities such as analysis of user requirements. By keeping out the non-important details, models allow people to focus on the essentials of complex problems. They represent the simplifications of reality; they are less complicated than reality (Gilbert and Troitzsch 2005). Models are an approximate representation of a more complex problem situation in the real world. The complexity of a real world problem situation may avoid the decision maker to reduce the chances of satisfying an exact representation of that problem. They need to find an acceptable balance between the manageability and accuracy of such problems. As a model comes out, a number of particular considerations will be there lead to some solutions. This would be the reason behind having several different models, each describing different angles of the real world. Models are several types range from simple (drawing models) to complex (computer simulation models). Base on researcher’s experience with modelling, time and data limitation, and issue of which model attend to address are the basics of choosing which model approaches are most useful.

Simulation is a “particular type of modelling” (Gilbert and Troitzsch 2005). It presents a new way of thinking about social and economic phenomenon. Simulations like statistical models have inputs entered by the modeller and output, which is the result of running the simulation. Usually, inputs are attributes used to close up the simulation with some specific social settings, while outputs are the model behaviours through the time. The output data can then be compared to the actual data (data collected from reality: survey data or historical data) to check how similar is the model to the real world. Simulations may be the best way to be used as “a method of theory development” (Gilbert and Troitzsch 2005) and testing for observing the behaviours (patterns) of a theory just by using a computer program. As a simulation is a particular type of modelling, it has the same purposes of modelling. One of the purposes of using simulations is for prediction. By using the model to look into the future if, the modeller could develop the model that truly reproduced some behaviours dynamics. For instance, simulating the alarming system of a building and test it to predict the weakness of the system rather than making a fire in the building then check it if it is working well or not. Developing new tools that might substitute for human abilities is another purpose of using simulations. For example, the expert systems (which might require experts) non-experts can use them to execute diagnoses. All of these mentioned purposes prescribe modellers and researchers to justify and declare their assumptions and theories regarding the system’s major components and the system itself (Kohler and Gumerman 2000).

The increasing of computers power and their storage capacity of these instruments has made a rapid revolution in using individual-level simulations or agent-based simulation modelling (Castle and Crooks 2006).

2.1.1. Agent-based Modelling

Agent-based modelling since 1990’s becomes the powerful new way to design, implement and analyze complex systems(Jennings, Sycara *et al.* 1998). Life is increasingly complex and need such systems to provide better understanding for its complexity. Agent-based models (ABM), Agent-based simulation

(MAS) Multi agent simulation model (MASM) or Individual-based model (IBM) are the most widely used abbreviations. The authors use them to refer to agent-based simulation models. The building blocks of these kinds of models are agents. These models consist of a set of agents that related to each others with a set of relationships using a framework to simulate agents behaviour and interactions.

Base on (Gilbert and Troitzsch 2005) there is no universal agreement on the definition of the agents. The term is used usually to represent "self-contained" programs that can control their behaviour and actions based on their perspective of their environment where they interact (Huhns and Singh 1998; Gilbert and Troitzsch 2005). (Russell and Norvig 2003) stated that the concept of "agent" is a tool for analyzing a system, but no absolute classification said a system entity (software, individual, etc) could be agent or non-agent.

Agents are of many properties. (Wooldridge and Jennings 1995; Epstein and Axtell 1996; Epstein 1999; Gilbert and Troitzsch 2005; Macal and North 2005; 2006) stated the following properties:

- *Autonomy:* agents are directing themselves in their actions without having others to govern them. They are driven by a need to achieve some internal goals. No "top-down" control over agents behaviour.
- *Reactivity:* agents are able to discover their environment and response to it.
- *Social ability/Interactivity:* agents are interacting with other agents using computer language rather than natural languages. In addition, they have the ability to interact with their environment.
- *Pro-activity:* agents are able to take the lead, and engage in goal-directed behaviours.
- *Heterogeneity:* "The notion of mean-individuals is redundant; agents permit the development of autonomous individuals. Groups of agents can exist, but they are spawned from the bottom-up, amalgamations of similar autonomous individuals".
- *Mobility*: is one of agent's useful features. They can move through the model space.
- *Learning:* agents can be designed to change their ideas and learning at the individual level or at the population level. They also have belief, desires, motivate and even emotions (Gilbert and Troitzsch 2005).
- *Explicit Space:* events occur on an explicit space, which may be a dynamical social network, or a landscape, GIS, or grid.

Agents need attributes to allow them simulate the phenomena they created for and produce results near to the real world data. Base on the purpose of designing they need some knowledge to base their action on. Some information should be provided to them as initial values then they will response and act. Their response may make them change their belief. By giving them a set of beliefs, they may extract further information from them. For instance, giving them the information "contacting with infected agent lead to illness" this may infer them to avoid contact/interact with the sick other agent to keep himself healthy.

The social ability of an agent made them to learn about interrelationships between other agents in their environment. For example, a parent agent may learn from his child agent who is the child's friend so that may make him becomes the friend of his child's friend parent. This is what called "social model" of an agent. (Gilbert and Troitzsch 2005) argued that agent's models "models are quite different from the simulation model that the researcher builds; agent models are built by the agents themselves while the simulation runs."

In order to let agents achieve their goals, they need a way to determine their behaviour that help them to satisfy their goals. One of the problems that the designer has is how to make agents know their own goals, also the way of deciding among several goals, which is the most important and relevant to the phenomena in hand. Having such engine "planning" is a solution of these problems. An agent may be programmed in a very straightforward to get that goal. For instance, an agent might be programmed that if there is an epidemic in the city they should go to hospital to get antibiotic to prevent them from infecting with the disease. It would be the matter “if you find this do that” however, agents programmed to do more complex planning.

Planning is a reverse process. Agent’s planning required to work backwards from a specific goal state (Gilbert and Troitzsch 2005) concluding what action result in that goal, what state might be needed before that event can be executed and so on until the designer gets back to the current state of the agent.

For example, in order to prevent an agent from infecting with the disease (this is the goal), taking antibiotic, going to the hospital, an epidemic should occur in the city, agent is susceptible and might got the disease (the current state of the agent). Even though there are existed some very sophisticated planners built by AI researchers but most human actions and reactions to a particular status is driven by routine reaction.

Interaction feature of agents gives them the property of being socially. All MAS’s include some type of interaction between the model agents or between the model agents and environment. This is the language between agents. This interaction may be of form of passing specific information from one to another or even an event occur with another agent lead to threat the agent with a specific information. For instance, an agent’s child friend may not go to the school that give him information of illness. Such communication should be modelled using a specific language for passing information. However, modelling human language lasts an area of debate and difficult to achieve.

An ABM has three components framed in such called “production system” (Gilbert and Troitzsch 2005). This system consists of three components: a set of rules, a working memory and a rule interpreter. Set of rules consists of “event-action” which is simply said when an event should occur then what should be the action. In order to make these roles occur the agent needs a working memory, which will help to store the information and facts. An event occur because a condition becomes true, then the action should response to the new event. The memory will help to store information about where or how the response should be. These rules (even-action) need an interpreter to check each role in turn. Checking whether conditions met then should be an action exists. The action of interpreter when the condition parts of more than one rule is satisfied is needed to be decided by the designer. The rule interpreter read the memory then applies what is stored there. If the memory contents changed during the previous action, the interpreter read the new values in the memory. For this case, the designer should be aware of specifying actions that effect the environment in a way that the agent would sense or directly affect agent's memory.

Agents are located in an environment in all ABMs. The environment structure depends on the model purpose of designing. However, if the agents are representing people in reality then the environment should provide them with a spatial context. The agents may move around the environment. Not only over a spatial space, agents can move but also they may through a network of nodes and links. The moment agents enter the environment they should be provided with a sensor to discover their neighbourhood environment. Model environment gives the agents the media, which they communicate through exchange messages, or may provide agents with messages by itself.

One of the ABM's components is the "time". Time is one of the decisions that the designer should make about the order of giving agents computation time. However, it is assumed to be operating in parallel among simulation's agents. The model schedule should be described because it deals with the order of the processes and updating the state variables (Grimm, Berger *et al.* 2006). It is the way of how the time should be modelled in the simulation.

Castle and Crooks (2006) claimed three main advantages of ABM's over the traditional models. These advantages are: 1) Captures the phenomena arising; 2) for the study of a particular system, AMB provides a natural environment; and, 3) the flexibility of the ABM particularly in simulating systems related to the development of geospatial models.

ABMs were considered an attractive alternative approach to deal with complex dynamic systems such as social phenomenon. These emerging phenomenon, which results from the interactions of individuals could be captured and analyzed by simulating a set of agents within the phenomenon environment. From the bottom up, agent-based modelling generates emergent phenomenon (Bonabeau 2002). However, it should be clear that individual behaviours are complex, and describing such system with the equations for instance becomes intractable so that ABM let modellers to think out of the traditional box. It can describe what is actually happening in reality. Epstein and Axtell (1996) stated that ABM may "change the way we think about explanation in the social sciences. What constitutes an explanation of an observed social phenomenon? Perhaps one day, people will interpret the question, 'Can you explain it?' as asking 'Can you grow it?''.

For describing and simulating a system composed of behavioural entities, ABM provides a most natural environment. It makes the model be closer to reality, explore the systems efficiency such as the ABMs which had been used to simulate the marketing models or the ones used for the policy makers to check the output of their decisions and how far they are efficient.

Particularly in the geospatial field and simulating dynamic spatial systems, ABM provides integration the model with the GIS. This integration allows the modeller to incorporate detailed real environment data, which are the representation of geography or space to simulate agents' behaviours and movement constrained by such data. In addition to provide tools for analyzing and simulating, the integration provides visualization the results in both 2D or 3D environment. Furthermore, real-time GIS data, which is crucial element in simulating and visualizing any phenomenon unfolding in real-time can be included. For the properties agents have, ABMs are considered the best choice to simulating social phenomenon and real world complex problems such as simulating land-use changes, epidemics.

2.1.2. Object Modelling

To transfer agents from the conceptual of their design to a computer program, object-oriented programming language (OOP) such as C/C++, Java etc. is used. It is the best way to improve the quality of the programming (Gilbert and Terna 2000). OOP provides the agent-based modelling with data structures that increase the efficiency of the model, enable the modeller to communicate model details, avoid difficulties of in modifying the programming code and allow other users to replicate the results. It provides the simple translation of the problem that needed to be simulated into a set of agents and events.

An object is a piece of code, which holds the data (state) and the methods (behaviour) that operate on this data and they provide the primary mechanism for object-to-object communication. The data for the object is similar to the attribute of the agent, and the methods of an object are similar to the internal state for the agents. With the use of high-level tools such as RepastS or Swarm, events are treated as objects. Scheduling these objects "in time-sensitive widgets" (Gilbert and Terna 2000) such as action groups.

OOP is used to design and build software agent systems(Gilbert and Troitzsch 2005). Methods in OOP can represent the internal state of agents as been stated. It includes agent's working memory and rules. They can be considered as implementer the rule interpreter. The structure of OOP is: template (classes) specifies the composition of the object which is known as instance. It includes concepts of inheritance, encapsulation and polymorphism. Inheritance means that assume there are two classes called root and leaf, leaf can inherit the root's attribute and methods then root is considered as superclass and leaf is the subclass of root. Polymorphism is simply meant that the subclass can have a method with a specific name that may exist with the same name in the superclass but it is possible that it does some other operation. Here, the subclass can use both methods and it will give the idea that subclass has a method does different operation. Finally, encapsulation means that requiring all interaction to be done through the methods of object and hiding internal state(JavaSun 2010).

2.2. Modelling epidemics

Subjects that have the direct relationship to human's life had been paid a lot of attention from the scientist. They provided tools and principles to represent better understanding and perspective to such issues. Epidemics are one of the most related subjects to human's life. They considered being an attractive field of research since long time ago. Epidemiological scientists did their best to model this phenomenon and help the policy makers by providing them with accurate predictions and tools to made decisions that keep human life safe. The better understood the diffusion of diseases the more efficiency the reaction can be (Claude, Perrin *et al.* 2009). Epidemics have been modelled with different types of models: mathematical- stochastic, network-based and computational approaches (both cellular automata and agent-based modelling) since past century.

The mathematical models assumed homogeneity in the population and it divided the population into groups of sub-population based on their infection status. SI models which refers to the two groups of the population (Susceptible and Infected). They assumed that every individual in the population has contact with every individual (Anderson and May 1979; Bonten, Austin *et al.* 2001). These models followed by some others all depended and referred to the classification of the total population such as SIR, SEIR, SIS, and SIRS. The “E” means the exposure, “R” recovered or removed based on the disease type. However, in reality these assumptions cannot be achieved. The human population is heterogeneous and individuals in the population are not in contact with all the other. Furthermore, it is impossible to discriminate on who infected and on individual responses (Claude , Perrin *et al.* 2009; Perez and Dragicevic 2009).

The second approach, which is the network-based approach, is used widely in simulating the disease spread. This approach depends on the social interaction between individuals. It represents the population's social networks. The approach assumed that the disease is spread along the edge which connecting two nodes(Eames and Keeling 2002; Toroczkai and Guclu 2007; Eames, Read *et al.* 2009). The nodes represent the individuals and the relationship/contact between them is the edge. For simulating diseases, which required social contacts to spread such as HIV/AIDS the network-based models, are very useful (Eames and Keeling 2002; Claude, Perrin *et al.* 2009). However, Claude, Perrin *et al* (2009) argued that for several layers of social interaction or casual contact between strangers in the crowded area network-based approach cannot account.

Agent-based modelling is one of these models that have been used to simulate the spread of the disease within people. ABMs are differing than traditional models not only in leading to better understanding the process of disease propagation but with their efficiency to simulate the heterogeneity of the population. It takes in account the temporal and spatial aspects of the diffusion of a disease. In addition, ABM's provide the individual contact and behaviour level of data. Furthermore, it provides tools to build and test theories related to diffusion process, test control measurements such as vaccination, design and analyze surveys of the epidemiological (Yang, Atkinson *et al.* 2008).

It is important to understand the process of disease transfer at the level of individuals (Yang and Atkinson 2007). This understanding is crucial for building the simulation model for the disease. ABMs were being used to simulate different types of epidemics such as influenza or avian flu (Ferguson, Cummings *et al.* 2006; Tuckel, Sassler *et al.* 2006; Markel, Lipman *et al.* 2007; Yang and Atkinson 2007; Dalton, Durrheim *et al.* 2008; Lee, Brown *et al.* 2009; Perez and Dragicevic 2009), smallpox (Epstein, Cummings *et al.* 2004; Barrett, Eubakn *et al.* 2005), HIV/AIDS (Teweldemedhin, Marwala *et al.* 2004). Usually, these models based on the interaction and the contacts between agents (individuals). The contact might be face-to-face or skin-to-skin or both of them to made the disease spread from infectious agent to susceptible agents. It could be transmitted via vector such as blood-borne diseases malaria (de Vries ; Gu and Novak 2009)

The disease ABMs are variance than each other in the way of simulating even if they are dealing with the same phenomenon (the same disease) and believe on the same principles. For instance, for both the above smallpox ABMs, they were dealing with spread of smallpox from different perspective and presented their model different from each other. Epstein and Cummings *et al.* (2004) used the grid cell to represent the population living in two cities, while Barret and Eubank *et al.* (2005) used the virtual city model (EpiSims) to represent the Chicago city environment. These differences in simulating such phenomenon is a very healthy result and proof the flexibility and efficiency of ABM's in such fields of studying. it is good to state that the models above could be termed as individual-based models (IBMIs as well as they are known as ABMs(Bian 2004).

IBM have the ability to explicitly describe the differences between individuals based on their attributes, which have direct affect the process of disease diffusion such as physical (e.g. immunity level), social (e.g. friends), economic (e.g. workers or non-workers), and environmental characteristics (Yang and Atkinson 2007)

Among all of these agent-based modelling and of course there were more we did not mentioned here, ISTAM model which is presented by Yang, Atkinson *et al.* (2007;2008) was the most attractive and interesting to start from in building this research's disease model. It provides a new way of building a dynamic contact network based on the individuals' activity patterns. The dynamic of the contact network means the individuals contact change over time the order of these contacts affect on the disease diffusion. The following section would provide a brief overview about this model and the principles presented.

2.3. ISTAM model

The ISTAM model (Individual Space-Time Activity-based Model) is a disease IBM model developed by Yang and Atkinson (2007) integrating "the contact pattern of individuals, an infectious disease process model, and a stochastic infection model by simulating ABs". It is time discrete stochastic model. Its implementation is based on Repast, java and UML. ISTAM consists of three objects: the infectious diseases, individuals and activity bundles.

For the infectious disease, they defined thee parameters to describe the applied infectious disease: effective contact, disease severity and evolution of the disease within the host. The disease diffusion in ISTAM depends on the distance between infectious and susceptible individuals if there are no obstacles between them as well as directly related to the duration of contact.

For an individual in ISTAM's model, s/he has three groups of attributes: static, dynamic and intelligent attributes. The attributes including demographic information, health level activity pattern that are not changed during the simulation and the reaction attributes to the current situation.

In ISTAM the individual space-time model is described at two levels: between activity bundle and within activity bundle.

Activity bundles or as in ISTAM known as AB are unites of space. In ISTAM space where most of individuals' interactions occur is divided into number of activity bundles. AB could be a room, a whole building or a building complex. They can be classified into different groups according to the human activity type and the function of the space. The parameters used to express an AB are the space-time dynamics of individual within the AB and the geometry of the space such as it is size and spatial layout. Movement within AB has been considered in ISTAM; in some AB, individuals remain static during the simulation time while in some other AB movement is required.

The contact network in ISTAM is generated from individuals space-time dynamics constrained both by the physical condition of the space and the individuals' activities. At the level of between ABs and within ABs the activity pattern, spatial effects and contact network are considered. At the level of between ABs the individuals are visiting locations near their current location more likely. their activity patterns can reflect this expectation. On the other hand, at the within ABs level, the effective contact concept is applied.

ISTAM was performed to simulate a hypothetical influenza epidemic at the University of Southampton. The first year of undergraduate students was the ISTAM population whom their data had been collected via a questionnaire survey. The questionnaire was consists of three parts: personal information, time point and activity preferences and durations. The information of this survey was used to build activity patterns per individual.

Total population was 4000 individuals belonged to 6 accommodation houses and 26 schools. No weekends and holidays were simulated. They were ignored. The structure of the model environment was estimated from real world data. Twelve types of AB were designed to represent the real environment. Among 315 activity patterns, one activity pattern was chosen randomly for each student.

ISTAM had been applied in the level of city. Eemnes, a city located in the north of the province of Utrecht in the Netherlands was the case study of ISTAM that time. The city population equals 9000 inhabitants and the city area equals 3362 ha. The city has been chosen because it is limiting exchange of individuals with the outside of city and not located in the direct vicinity of larger cities. To simulate ISTAM with this city, four data sources were used: activity survey data (ACT), synthesized household data (HH), land use data, and PC6 statistical data.

The ACT data were supposed to be representative for people residing within Eemnes. To select activity data ISTAM strongly depended on this dataset (i.e. ACT). The environment data were selected using the land use data, which contained the number of individuals, engaged in every vocation. The HH data were grid based therefore; all households within all grid cells of the city were pooled together then allocated each household with its family member to the PC6 of the city.

The time unit was set and one day was divided into number of unites. These time unites were assigned with specific activity type during the simulation of a person's daily activity. Study time, work time and sleeping time were assigned to individuals as personal attributes during the generation of whole the population and were fixed for the daily activity of one-week day. The remaining time unites were assigned to activity types based on the probability weights from the assigned patterns related to the socio-demographic data.

Houses, workplaces and schools were fixed to every person in the simulation from the beginning. Randomly the other locations were selected usually the spatially closest one. Individuals can move between different locations at every time step.

The ISTAM simulation results can be extracted at any level. Therefore, it is calibration and validation had great chance to be provided. The ISTAM calibrated two of contact indices: contact frequency index and contact intimacy index. However, until the time of publication of their model validation was under way by comparison of the simulated number of individuals present in different types of AB over time with related research results.

2.4. Concepts of this disease modeling

2.3.1. Pertussis childhood disease

Pertussis or Whooping cough is a communicable disease of human, occurs due to Bordetella Pertussis bacteria that arises through transmission of the bacteria in airborne droplets from infected person to a susceptible person; either directly via human contacts or indirectly through the inanimate environment. Pertussis is a highly contagious infection pathogen of respiratory tract particularly in children. Globally, the disease caused a significant part of childhood morbidity, discriminated by continuous acute paroxysmal cough. This characteristic is distinguishing the disease as whooping cough.

Before the introduction of vaccination, pertussis was one of the main reasons behind infants' incidence and death. It was causing more mortalities than all other communicable diseases of childhood(Boven, Hester E. de Melker *et al.* 2000). After the introduction of vaccination, the illness and deaths because of pertussis are sharply decreased. Notwithstanding and based on WHO reports, whooping cough continuous to put millstone, where there are still 50 million incidents of Pertussis (90% of them are in developing countries) and 300,000 deaths yearly, most of them among infants (WHO 2007).

Over the last twenty years of the last century, infections with Pertussis were reported in increased rates in several countries. In UK, Japan, and Canada, the increasing of pertussis cases had been reported as a result of using ineffective vaccine (Edwards and Halasa 2004) while in France and USA investigator had suggested that Pertussis is increased in adolescents and adults (Güriş, Peter M. Strebel *et al.* 1999; Gilberg, Elisabeth Njamkepo *et al.* 2002). However, these countries are considered to be vaccinated population with a high level of coverage. The Netherland is among these developed countries with high degree of vaccination (96% of the children have at least three doses of vaccine) but Pertussis shows an increasing tendency (Bentsi-Enchill, Scott A. Halperin *et al.* 1997; Boven, Hester E. de Melker *et al.* 2000).

Pertussis is a childhood disease and primarily affecting children of less than 10 years old however during the last three decades the proportion of cases involving adolescents and adults have increased and across all age groups (Boven, Hester E. de Melker *et al.* 2000; Gilberg, Elisabeth Njamkepo *et al.* 2002; Hoey 2003; Edwards and Halasa 2004). For instance, 50% of reported cases in United States were people 10 years of age and older (Hoey 2003). Moreover, the disease seemed to be as a contagious among adults as it was among nonvaccinated children. Based on the previous and the number of cases that had been reported and registered of infected adults also the demographic shift of the disease gave the impression that adults and adolescents are major reservoir of pertussis in the community. Pertussis is very infectious with high secondary attack rates in households. Wendelboe et al showed that parents, siblings and nonhousehold close contacts are important source of pertussis diffusion to infants. They provided robust evidence that household members are in charge of 73% - 82% of pertussis diffusion to infants(Wendelboe, Njamkepo *et al.* 2007).

One of the reasons behind this demographic shift to adults and adolescents is because the immunity following vaccination decreased and waned (Bentsi-Enchill, Scott A. Halperin *et al.* 1997; Güriş, Peter M. Strebel *et al.* 1999; Boven, Hester E. de Melker *et al.* 2000; Gilberg, Elisabeth Njamkepo *et al.* 2002; Hoey 2003; Edwards and Halasa 2004; Wendelboe, Njamkepo *et al.* 2007). For this reason, it had been suggested that in order to protect infants is to increase community immunity by vaccinating close contacts. It had been recommended a booster dose of vaccine for parents and health care workers who are in contact with young children (Forsyth, Nagai *et al.* 2005).

Pertussis is transmitted from person to another during coughing or sneezing and producing respiratory droplets. The Chinese called pertussis with the “cough of 100 days” due to the prolonged, dry cough. Usually, pertussis can be difficult to diagnose (Hoey 2003) and that because early symptoms of whooping cough may occur like the common cold. Pertussis incubation period is last for a week to 10 days and is characterized by the insidious onset of cold (runny nose), sneezing, low-grade fever and a mild, accidental cough as the common cold nothing special for pertussis. Then gradually the cough becomes more sever.

After 1-2 weeks, the second stage or period starts. Usually, paroxysmal stage is the stage where the diagnosis of whooping cough is suspected. In this stage, cough is characteristic followed by whoop sound, vomiting, or both. Normally, clinical severity varies widely among different age groups and vaccinated individuals (Hoey 2003). For instance, adults rarely have the classic “whoop sound”. This whoop sound followed coughing gave the disease the popular name of “whooping Cough”. The paroxysmal stage is followed by two months of coughing episodes.

Communicability period of pertussis begins from the onset of symptoms to 3 weeks after onset of cough. Transmission of pertussis is mostly occur during catarrhal stage (probably from early in this period of illness) to approximately 3 weeks before sever coughing spells. Through the direct contact with airborne droplets broadcasted to the air during cough or sneezing by the infected individual the transmission commonly occurs. Rarely, transmission occurs through contact with contaminated surfaces around.

The last stage of the illness before the recovery is the "convalescent stage". Cough may continue for many weeks during convalescence. It takes around 6 to 12 weeks. During this period, coughing paroxysms and intensity gradually decrease.

As been mentioned previously, the severity of pertussis is varying widely among different age groups. Infants aged < 12 months are more likely to have severed pertussis and suffer from pertussis-related deaths or severe complications than older age groups, partially vaccinated individuals. Adolescents and adults may become infected with pertussis but within milder disease. The infection may be asymptomatic, milder, or even classic pertussis with persistent cough lasting for more than 7 days. Even with milder infection but the transmission of the disease to other susceptible individuals. These adolescents and adults are important source to diffuse pertussis to infants and small children.

Worldwide, more than 50 million cases around the world with more than 250 000 deaths (WHO 2007). Pertussis is human disease. The important reservoir for the disease is adults and adolescents and usually they are the source of infecting infants. Pertussis has no special seasonal pattern, but it may increase during summer and autumn. Pertussis is highly communicable and it has secondary attack rate of 80% among susceptible household contacts. Pertussis epidemic occur every 3-5 years. Transmission occurs by close contact. The close contact defined as anyone has had direct contact with airborne droplets polluted with infected individual's respiratory tract.

2.3.4. Social Interaction and disease spreading

Understanding ways and patterns of diseases spread via population lead to better controlling the diffusion of these diseases and preventing future epidemics by developing efficient strategies (Toroczkai and Guclu 2007; Read, Eames *et al.* 2008). This understanding can be achieved by determining and defining which characteristics are most important in dictating the patterns theses diseases are spread. These characteristics are proceeded from the demographics and connectivity of people existing in the same society. For instance, Grenfell *et al.* had proved that a shift in epidemic periodicity from biennial to annual in a population above a defined threshold could occur because of increasing of birthrate (Grenfell, Bjornstad *et al.* 2002). In addition, the high density of population and its people interaction lead to significant risk of disease spreading and thus on epidemics.

Human beings are living, working, interacting etc. together in a community. They have complex relations and day by day, the complexity is increased. This increasing complexity of humans' relation and transport infrastructure are the key factors of disease epidemics (Colizza, Barthélemy *et al.* 2007). Via this complex system of humans relations infectious disease transfer from person to another and from place to another infecting new persons and places. This complexity as well as the social structure adds heterogeneity to population, which have been argued that they have a large effect on diseases spread (Galvani and May 2005). In particular and because of the negative impact of epidemics on a population, development of mitigation and prevention methods gain great importance (Toroczkai and Guclu 2007).

For the spread of diseases, not only the distance between people determines the risk of infection but the interaction of people. Human's interaction pattern has important implications for the spread of infectious diseases. (Keeling and Eames) have argued that the individual with many contacts is at risk of infection, the time infected, the disease then can be transferred to many others (Keeling and Eames 2005). This risk considered to be in high level for the diseases such as HIV where the majorities are in long-term relationships then identifying the 'core groups' of individuals with large number interactions helps to maintain of control policies for these type of diseases (Hethcote and Yorke 1984). While for airborne infectious disease such as SARS (Severe Acute Respiratory Syndrome) the significant proportion of all infections were referable to super-spreaders (super-spreader is an individual who is infected and may generate many more than the average number of secondary cases (Riley, Fraser *et al.* 2003).

The social interaction means every type of contacts two individuals have together when their space-time intersect. These contacts may be casual contact (face-to-face or conversational) or close contact (skin-to-skin or physical). Different diseases seek different levels of interactions to effect transition (Read, Eames *et al.* 2008). For instance, smallpox is thought to require very close contact between an infected and a susceptible two individuals to lead to infection, while measles, pertussis, and influenza are considered to transmit easier via airborne droplets; therefore, they may only seek a conversational proximity between people to transfer (Epstein, Cummings *et al.* 2004; Read, Eames *et al.* 2008). In addition, infectious sexual disease such as HIV/AIDS required sexual (physical) contact to be transferred and conversational or proximity contact will not lead to diffusion. However, it had been argued that interactions between individuals who are the transmitter of the airborne diseases are harder to define and appear at greater frequency than sexual contacts (Eames and Keeling 2002). In a detailed diary-based survey of 49 adults presented by (Read, Eames *et al.* 2008) showed that per a day individuals have approximately seven times more conversational contacts than physical contacts.

The social interaction is differing from a culture to another. Even in the same society people with different ethnics, their social interactions are not the same. In a case study presented by (Tuckel, Sassler *et al.* 2006) who used GIS to measure and track the incidence and the spread of influenza pandemic of 1918 in Hartford, they argued that epidemic is a discrete event occurs within the confines of one city and affects various subgroups in the population. The major point of differentiation of these subgroups was ethnicity, which overlapped with geography. For safe probably, they assumed that immigrants to USA from southern or eastern European who lived in areas not heavily populated with individuals from their same nationality background had a lot of contact with their fellow immigrants occupying the same job settings and keeping close social ties (Tuckel, Sassler *et al.* 2006). They argued that in addition to factors that are biologically based, such as age, social factors also should be based.

The social interaction is different from an age group of individuals to another. It is assumed that children in general are more interaction than adults are. They have very close interaction in schools for instance, for there close interaction in cases of epidemics occur the first think the policy makers did is school closure (Markel, Lipman *et al.* 2007; Dalton, Durrheim *et al.* 2008; Lee, Brown *et al.* 2009). During epidemics, children transmit the disease to one another then later to their parent and households members. Consequently, closing schools will eliminate one of the important transmission sources. However, it might be not efficient decision due to the significant effects on economic and social disturbance to the local society by paying additional burden on parents(Lee, Brown *et al.* 2009). Parent forced to stay at home taking care of their children when schools are closed that may lead them to lose their jobs, which could be additional problem for the families with low income.

The individuals' social interactions are constrained by both the space and time, and the intimacy and social context (Yang, Atkinson *et al.* 2008). Social interactions can be characterized in many different ways. This characterizing is including distance of contact, type of contact, frequency of contact, duration of contact etc.

Distance of contact is the physical distance between two individuals in contact with each other. Usually it is assumed that two individuals come into contacts who are physically adjacent (Yang and

Atkinson 2007). They perhaps have a great chance to encounter the people who are spatially adjacent to them. However, individuals can be in contact with other individuals who are not physically adjacent them; it might be via telephone, or email. Of course, this is not countable for the cases of disease spread. The geographical distance alone is enough to measure infection risk between individuals (Bian 2004)

The frequency of interaction means that how often or how regular are individuals interact (Read, Eames et al. 2008) since it is not adequate only to measure the number of contacts each individual made but also the frequency of these contacts. We denote this frequency of meeting or contacting by "weight" (Read, Eames et al. 2008). This weight helps to explore the impact of the heterogeneous contacts on the diffusion of the disease so the risk of individual infection can be estimated and a prevent strategy such as vaccination can be applied effectively or might be closing the building with most infected individual to control the outbreak of disease (Eames, Read et al. 2009). In another words, the weight will provide an avenue to assess the opportunity that infection will spread along the link of the interaction.

The duration of an interaction is the amount of the time that individuals are spending while interacting. This time is important, as its length will decide if it is enough for a disease to transfer from an individual to another.
With the type of interaction we distinguish between physical contact -skin-to-skin- or conversational contact without touch, casual contact or close contacts (Read, Eames et al. 2008). Individuals live, work and do life's activities together (i.e. the range to which they govern the network) that means the social interactions have importance to them therefore; it is probably closely connected to their role in disease spread, such ideas are directly relevant to epidemiology (Epstein, Cummings et al. 2004).

A large number of social interactions together are often referred to as social networks. Usually these two terms (Social interaction, Social networks) together were being related to spread of diseases. They provide a better understanding and wider perspectives to simulate the diffusion of diseases and epidemics.

2.3.5. Social Networks

The diffusion of infectious disease is considered as a crucial social problem. The civil infrastructures, human behaviour, and social networks all have an important role in initiating and controlling these epidemics. Studying the diffusion of disease on social networks recently has attracted considerable attention. They have more flexible and richer capability for representation of scenario (Newman 2002). This approach to modelling epidemics provides a new theoretical foundation for parameter estimation, data analyzing, forecasting and modelling (Connell, Dawson *et al.* 2009). The approach handles the structure of relationships and the interactions between individuals. The development of agent-based modelling infrastructure is a crucial step on its response to the epidemiological events and the realistic simulating of complex social network dynamics. Two parameters control the diffusion of diseases through population: the disease properties (infection mechanisms, symptom timing etc.) and the structure of the social network of the population (migration, contact rate, and social clustering). However, these social parameters are completely unrelated and can independently be studied (Connell, Dawson *et al.* 2009).

Social network models a system consists of individuals (nodes) and their connection relationships (ties). According to (Carley 2009) the social networks are " an area of computational modelling of increasing importance for assessing and understanding human social behaviour". They can be used to explore the issue of veridicality and simplicity.

Social networks importance (advantages). Dealing with social networks in modelling human behaviour and interactions is of crucial importance. Many reasons are behind this importance (Carley 2009). First, among all computational modelling fields, network science is the most developed because of the easily linkable to other types of models, availability of validated, documented, well understood, and meaningful metrics, toolkits, well-understood procedures for data collecting and analyzing. Second reason, network models have been used with concrete success to support decision making in fields such as counter-terrorism, corporate re-organization, and social policy. Third, they enable and constrain the behaviour to that extends of identifying the factors and supports course of action analysis by understanding the network in a group such as who interact with whom influences the spread of disease, what information is transmitted or learned and the flow of money. Forth, to create a network model it is a matter of minutes to be created. Any set of nodes, which have known links can be imported to any network analysis tool then visualized, its metrics calculated, the impact of its nodes or links removal or adding assessed. The social networks give the chance to think about the dynamics of meta-networks that connect the who (e.g. individuals), what (e.g. activities and events), where (e.g. a building), why (e.g. goals) and how (e.g. the type of contact). These dynamics should not be underestimated (Carley 2009).

2.5. Other Concepts in use

2.4.1. Synthetic population

Generally, micro data of a population is suppressed so that to keep confidentiality or uncompleted because of the high cost of its acquisition. Therefore, using the techniques of synthetic population are come up as an executable alternative to the collection of micro data(Beckman, Baggerly et al. 1996).
Synthetic population techniques are methods which take data of aggregated population in addition to sample population data as input, and the output would be a complete list of population individuals, each with associated attribute data(J. Ryan, H. Maoh et al. 2007).
Different methods may produce synthetic population, which all adjust to the input data, but differ in their quality(J. Ryan, H. Maoh et al. 2007). The crucial aim of developing and building a synthetic population is to create the requisite population as precisely and accurately as possible (Juliet Nakamya, Elke Moons et al. 2009) when compared to the realistic population.

Different methods may produce Synthetic population which all adjust to the input data, but differ in their quality. (J. Ryan, H. Maoh et al. 2007). A number of approaches exist for estimating spatially detailed microdata, including stratified sampling, geodemographic profiling, data fusion, data merging, iterative proportional fitting, reweighting, synthetic reconstruction and their combinations. The last two methods considered to be the main methods of creating the synthetic population (Huang and Williamson 2001/2)

The Synthetic Reconstruction Approach
This method is a sequential procedure, creating members of synthetic population one by one using a set of conditional probabilities. It is a Monte Carlo based approach.

The procedure begins by creating a set of household heads with attributes age, gender, marital status and spatial location specified by sampling from a known distribution. A next step might be specifying the economic activity to the sample based on another published tabulation to determine a condition probability of economic activity. The next step could be to estimate her/his career then if married, a spouse and potentially, child(ren) might be generated. The procedure will carry out for all the variables one likes to include in the synthetic population.

Here, the only source of population is the census, which provides a series of predetermined, separated, aggregated cross-tables such as age by sex, by marital status for individuals and tenure by ethnic group of head of household for households. It is almost always the case that these tables provide only partial information concerning the conditional chain probabilities to be derived. The required conditional probability can be estimated using iterative proportion fitting (IPF), a well-established technique for overcoming data shortfalls of this kind. Specifically, IPF can be used to estimate the full joint distribution.

One of the advantages of IPF is that any number of sets of constraints can be embedded within the procedure. Hence, through IPF, the data contained in separate tables may be linked together. In essence, synthetic reconstruction approach tries to reconstruct the original population in such a way that all known constraints (the counts represented in the census tables) are reproduced. One main advantage of the synthetic reconstruction approach is that its use of conditional probabilities allows data to be incorporated from the widest possible range of sources.

For synthetic reconstruction approach the main problems are: A stochastic process Monte Carlo sampling is subject to sampling error. This error is likely to be more significant for small area simulation where the sample sizes are small. The ordering of conditional probabilities. A certain amount of error is introduced in each stage, which may be contributed variously by Monte Carlo sampling, modelling assumptions, and data inconsistency.

Combinatorial optimization Approach

It is an alternative way of estimating small area micro populations through the reweighting of an existing large-scale micro data set. The case study is normally divided into mutually exclusive and exhaustive zones such as census tracts, enumeration areas or traffic analysis zones. Two forms of input data are needed. First, a micro sample from the whole population (survey micro data) where the members of the sample contain the attribute variables desired in the members of the synthetic population. Second, tabulations for a sub-set of the desired attribute variables are required, showing the distribution of those variables over the zones comprising the study area.

The process is iterative, starting from an initial set of households chosen randomly from the survey micro data sets, an assessment is made of the effects of randomly replacing one of the selected households with a fresh household from the survey data. If the replacement improves the fit, the households are swapped. Otherwise, the swap is not made. This process is repeated many times, with the aim of gradually improving the fit between the observed data (a set of pre-selected constraining tables) and the selected combination of survey data households.

1. Step1:- Obtain sample survey microdata and small area constraints.
2. Step 2: Randomly select two households from survey sample to act as an initial small-area microdata estimate.

3. Step 3: Tabulate selected households and calculate (absolute) difference from known small-area constraints.
4. Step 4: Randomly select one of selected households. Replace with another household selected at random from the survey sample, provided this leads to a reduced total absolute difference
5. Step 5: Repeat step 4 until no further reduction in total absolute difference is possible

The combinatorial optimization approach provides a novel solution to the problem of generating small area micro data, but there are several areas for further investigation and refinements:

- Selection criterion: The existing combinatorial optimisation model uses total absolute error (TAE) as the measure of fit during the iterative fitting process, but the fit of the final synthetic data to the known small area tables is evaluated based on a relative statistic (Z score).

- The existing combinatorial optimisation model selects households from the whole survey data.

- Table fitting sequence: The ordering of tables in the sequential fitting procedure is area-specific, and the target level of acceptance is table-specific.

Based on the data available for this project, the first approach would be used in the simulation. The reason of selecting the first method is due to the details, which are available within the data and the purpose of using the synthetic population in this model. Details about the method used and the details of data were addressed in chapter

2.3.2. Human activity patterns

The approaches of activity-based "describe which people pursue, at what locations, at what times and how these activities are scheduled, given the locations and attributes of the potential destinations, the state of the transport network, aspects of the institutional context, and their personal and household characteristics" (Ettema and Timmermans 1997). The patterns of daily activities determined by constrains of time and space. A sequence of activities is defining a path in space and time. Cullen and Godson (1975) assigned six different phenomenon that affect on activity patterns. First, activities are performed with particular objectives or goals in mind. They are structured consciously so that emerging behavioural patterns. Second, individuals act within a structures spatial setting. These setting include opening hours, locations, services and facilities, and the like. Third, priority of activities depend on their importance, involvement of other people, frequency etc. Fourth, activity constrains includes time, money, available transport etc. Fifth, some activities are special combinations will be consciously scheduled to optimally use space and/or time whereas some other combinations will be planned routine-like. Finally, different activities are with different degree of flexibility. An individual can change an appointment time and/or space more flexibly than its work.

The importance of daily activity patterns comes from their constitution the confrontation between the individuals behaviour and the physical system of urban (Ettema, Jong Kor de *et al.* 2005). It is through their daily activities the households/individuals are experiencing the externalities and respond to them. While their aggregated behaviours lead to system externalities and user levels such as pollution.

The principle of activity-based approach was able to attract researchers to consider it in simulating some aspects in the social that are related to human daily life. They have been considered in the studying of land-use transportations models that is dealing with the changes on the land use and humans interaction and affect of that change such as Predicting Urbanisation with Multi-Agents (PUMA) (Ettema, Jong Kor de *et al.* 2005), Land-Use Transport Interaction (LUTI) (Ettema, Arentze *et al.* 2009), modeling household activity travel behavior (Timmermans and Zhang 2009). Within disease spread they have been used in simulating the patterns of diffusion during epidemics such as (Bian 2004; Eubank, Guclu *et al.* 2004; Yang and Atkinson 2008; Yang, Atkinson *et al.* 2008).

3. Conceptual Design

This chapter is aimed to present and describe the conceptual design of the Pertussis model based on what have been showed in the previous chapter. The components are discussed according to three stages of the simulation process, the initialization and setup of the model (stage 1), the actual simulation model (stage 2) and the visualization of the output (stage 3). The stages of building the model are illustrated in the Fig.3.1 and as the following:

- **Phase One**: Creation of the population and consists of the following steps:
  - Build Synthetic Population (3.1.1).
  - Load the environment (3.1.2).
- **Phase Two**: the actual simulation model and consist of:
  - Daily activity Model (3.2.1).
  - Social model (3.2.2).
  - Disease model (3.3.3).

- **Phase Three**: this represents the last stages of the simulation in which the output model will be visualized as the social networks of agents.

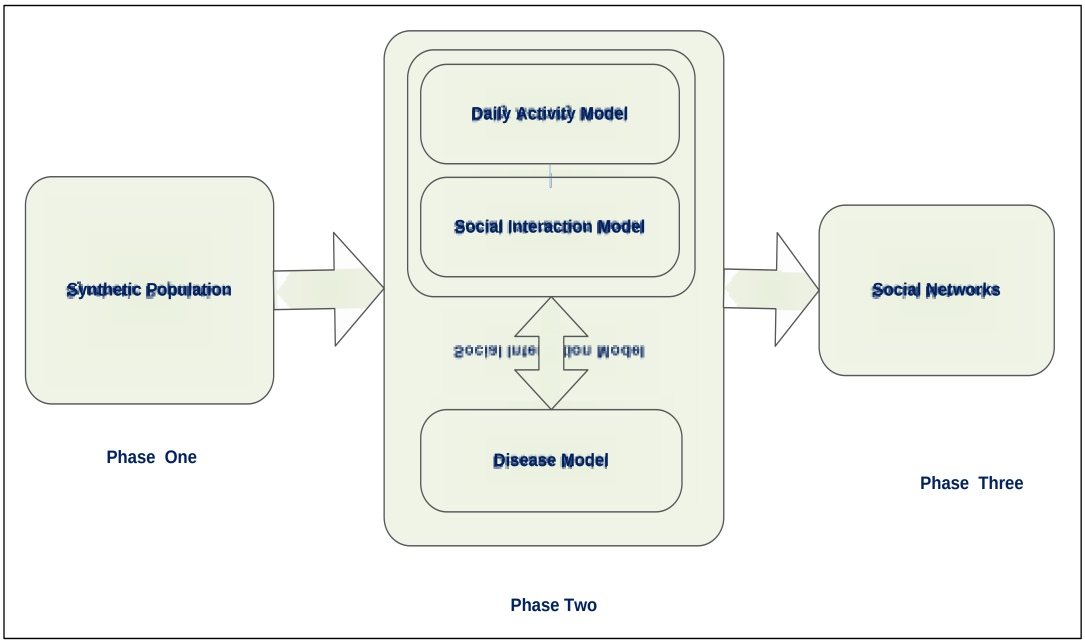


Figure 3.1: Pertussis model stages

3.1. Stage one: creation and distribution

For the simulation of the spread of a disease a population needs to be created that correctly reflects the true population of the area for which we are simulating, in this case the city of Enschede. The population should be realistic in composition according to age groups, marital status etc., but also according to their home location. For this simulation, we determine two steps in the creation of the population.

3.1.1. Synthetic population of Enschede (Agents)

The simulation of infectious disease diffusion requires the model to have a population that the disease will spread through it. Population consists of individuals who follow their daily activity patterns. These individuals have a list of associated attributes (e.g. age, gender, marital status, etc.). Typically, individuals are grouped into families and households.

The building block of this simulation is agents. An agent is a crucial part of the simulation system that senses the environment they would posed to in order to find the location of the next activity and move there. According to the rules and characteristics has will perform his/her own agenda of activities and interact with the other agents over time. As a result, the agent perhaps be influenced or affect on other agents while acting in the environment.

Agents are created as members of that they are members of a household living together in the same house) but can also be members of a larger family (aunts, uncles, grandparents etc) that are not living together. By using the synthetic reconstruction approach of creating synthetic population. This method will use the census data of Enschede to build families of different types (couple with children, single mother with children etc) and families of different sizes. Agents will obtain, estimate, assign their attributes values from Census data or the other datasets that have been used within the simulation such us the vaccination level type, work status etc.

Agents differ in terms of attributes that effect disease diffusion including physical (susceptible, infected, immune), social (age, marital status) and economic (worker, non-worker, full time work etc.) aspects.

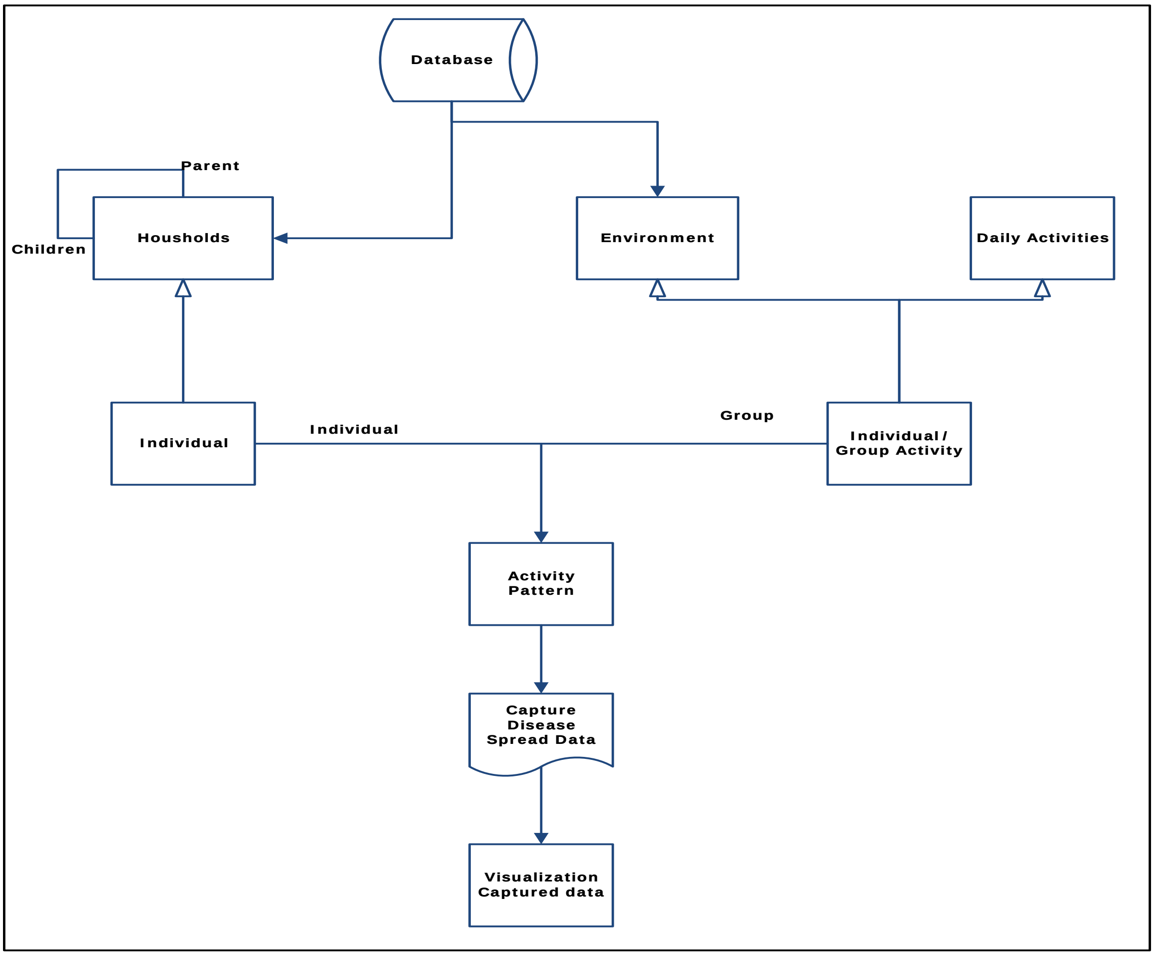


Figure 3.2 Abstract diagram of the model

In real live, some individuals are usually together, thus forming a group (Ellegård 1999a) such as family members, students in the same classes in school etc. These individuals have some common characteristics that are common among the group members such as the same house, which all members are living in, the same age value for the same class students and so on.

On the other hand, within a group, individuals have attributes that distinguish them from each other. For example, different individuals within the same family have a different age, students in the same class come from different houses, workers within the same department have different marital status etc.

These groups or engaging individuals within groups will help to determine the individuals' behaviour by the role they play in their group, and the roles they have within their groups (joint activities) are the elementary foundation for simulating the contacts between individuals (Yang, Atkinson et al. 2008). In addition, the model has group behaviour. The group behaviour is the activity that consist of sub groups of individuals that behave like one individual. Couple of individual who consist a group behave as only one individual. For instance, during weekend, the household members will follow their household head in participating the same activity with other groups with the same principles.

Individuals are in different types such as child or adult, worker, or student. These types have different activities and reaction to the disease if they or one of their group members gets infection.

Before creating synthetic population, some assumptions should be making for the purposes of simplification. These assumptions are:

- The population has a fixed size i.e. no new births or deaths during running the simulation will occur.
- Families with children, the maximum number of children are four.
- Single fathers with children, the maximum number of children are two.
- Single mothers with children, the maximum number of children are three.
- The range of parents’ age is considered to be in the range of (25 to 50) years old.
- Families with no children and singles with no children could be at any age starting from 17 years to 65+ years old based on the probabilities estimated from census data.

The next step is the distribution of these agents over the city space to give the model city shape consists of its people and features.

3.1.2. Model Environment and Spatial Distribution

Synthetic population needs to be spatially distributed over the city. As had been clarified previously, the agents are created based on the census data of Enschede. In this next step, families and their individuals should be connected to the spatial environment, which is represented by buildings. Each family should have a house and each individual has to look for his/her activity places and save their address in his/her memory to be used during achieving their activities.

It can be stated that building is the station where individuals spend a part of their daily time to do some activities. A building has a couple of attributes that would help the system during the simulation. The building type will illustrate the type of the building whether it is a house, a school, a workplace etc. Each of these placed has different activities and different type of interacting between individuals. In addition, each building has an address that would help directing individuals to the next activity place. Information that will be used to provide individuals with locations of different buildings and places and distributing them in a real way are from both the GIS vector dataset and the administrative dataset. Fig 3.3 is showing a part of these data and environment.

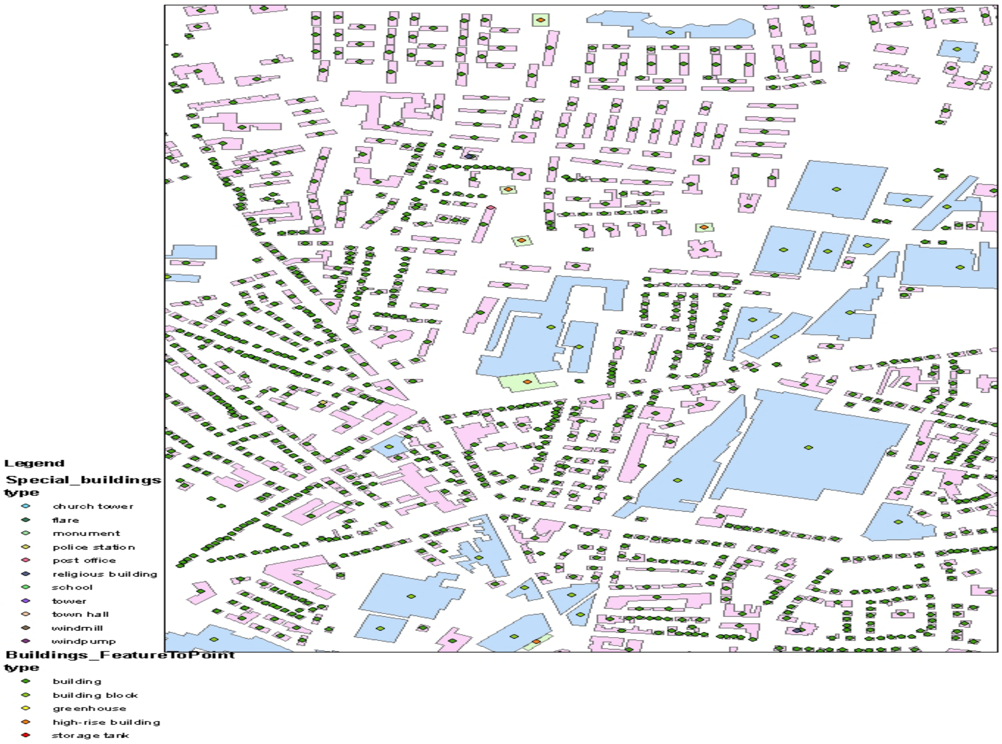


Figure 3.3: distribution over the city (polygons refer to the building shape)

The moment the simulation starts, all the households and their members are at homes. For the children the distance is the judge of choosing the activity stations. Even in reality, people choose nearest locations for their children activity for example, parents send their children to the school nearest to their house location. The model does not consider the school type or the sponsor of the school. In Netherlands as well as many countries, religious communion (private schools) or government (public schools) sponsors schools. Three types of elementary schools mainly exist in the Netherlands. These schools are catholic, protestant and public schools. Dutch people often used to send their children to schools, which they supposed to be from their same communion. “Some two-thirds of all Dutch children attend particular schools, of which the majority are Roman Catholic or Protestant,” (Oppen-Stuyt 2009). However, with the missing of these data the only role of choosing a school is the distance between the individuals’ house building and school building. The same role is applied for selecting the social places for the children’s social activities such as sport centre, BSO etc.

For the reason of simplifications, some assumptions have been made:

1. Agents type Child of age (0 – 3): they attend the baby care centre in the same neighbourhood and the rest of locations are with their families.
2. Agents type Child of age (4 – 12): they attend the nearest primary school (the shortest Euclidian distance between their house building and the school) with the condition that the building capacity still can accept agents otherwise should look for the second nearest one and again check the capacity of the next one and so on. For the other activities, again they should be in the same neighbourhood.
3. Agents type Child of age (13 – 18): they attend the high schools in any neighbourhood in the city. The other activities may be in the same neighbourhood.
4. Agents type Adult who are workers: these are distributed to different workplaces in the city with the consideration that a proportion of them may work somewhere outside the city.

5. Agents type Adult who is non-workers can attend any activity in their neighbourhood or other neighbourhoods.
6. The environment is static during the simulation, i.e. it will not changed by agents' activities.

Within these assumptions and based on the available data all agents will be distributed spatially in order to achieve their daily activities.

3.2. Two: Actual Simulation

The actual simulation model will start with this stage. This stage is simulating the individuals' part of life after they had been distributed spatially. In this stage individuals have to achieve their daily and weekly activities and interact with the other agents. During activities, infection and spread of disease may occur because of attending these activities which leads to a change in the normal daily pattern for the ill individual and the once who should take care of him/her (see fig. 3.4).

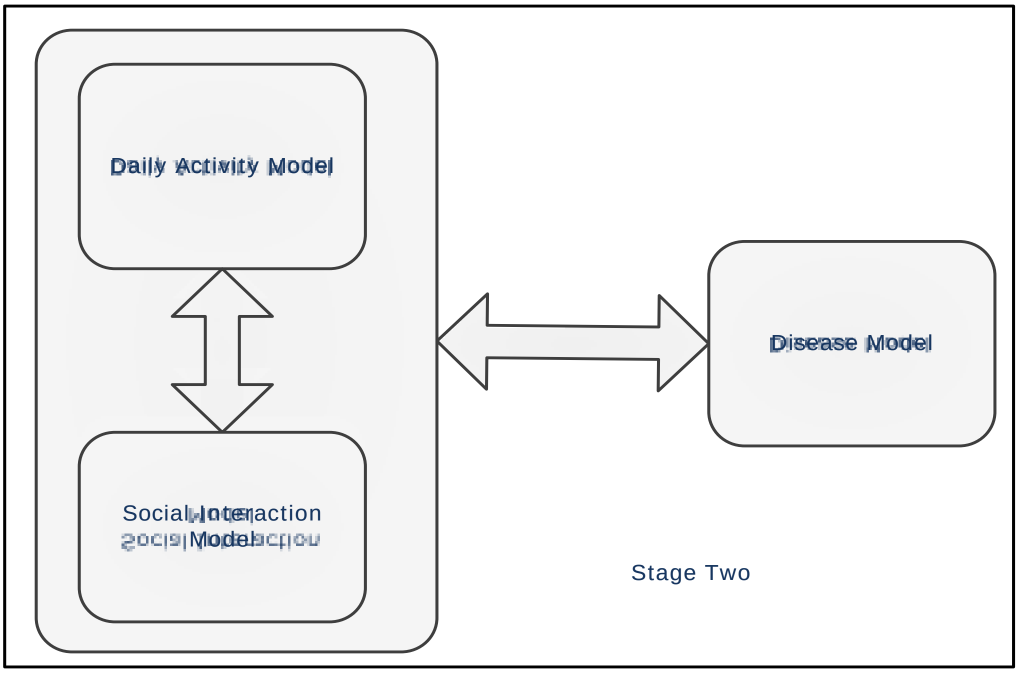


Figure 3.4: the relation between the stage two components

The following two sections will be described in more detail both: Daily Activity model and Social Interaction model, and the Disease model.

3.2.1. Agents' behaviour (Daily Activity Patterns)

While individuals are performing their activities, they may share the time and the space together and produce a group activity. A group is defined by its repeated meetings in time-space because of the social obligations of the member (Ellegård 1999b). These groups can be of the same age and goal like the group of a class in a school or they might be in different ages but has relationships together like the family group. An individual may have only one group every time like the member family who has her/his own activities only with her/his family members or may have multi groups like the individual who join several different activities with different individuals.

Every individual has to perform his or her activities during the time span of the simulation. Whenever the space-time of two individuals intersects (and they participate in the same group) that means they are sharing performing the activity. The activities are various in their durations of time. By the end of an activity an individual should change the current end activity to the next one based on their activity schedules. This procedure is continues until the end of the day.

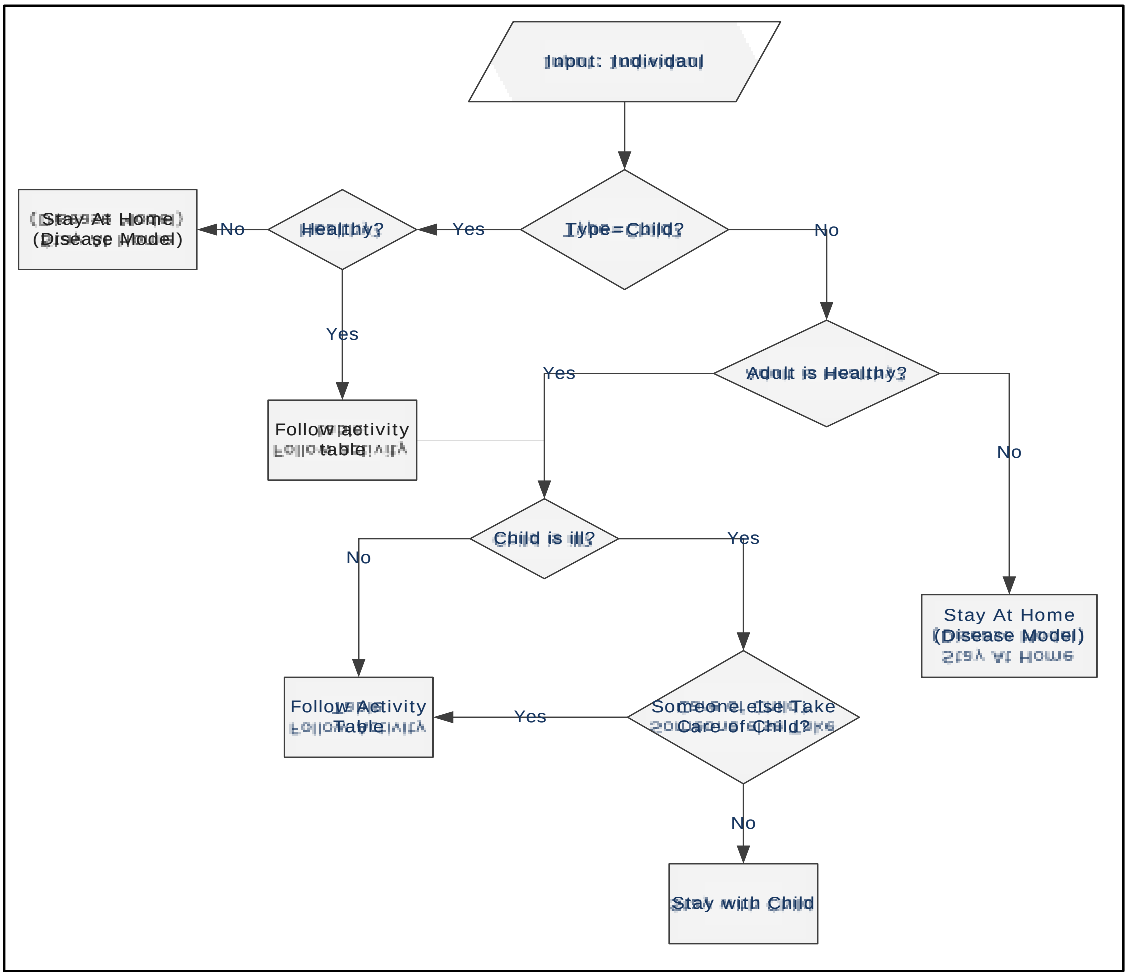


Figure 3.5: Explain individuals attending activity

Within a week some of activities appear repeatedly (like for instance going to school, work from Monday to Friday), while some others may occur once to twice a week (like attending a music class or visiting a friend). The daily activity model produces an activity pattern table per individual. It will have a forma like the following table.

WEEK1:MON	ACTIVITY 1	LOCATION	TIME	DURATION	...	ACTIVITY N	LOCATION	TIME	DURATION
Week1: Tues									
Week1:Wed									
...									
Week N: Fri									
Week N: Sat									

Table 3.1 Individual Activity table: everyday in a week in whole the simulation

The first week of the simulation will be the base for the next coming weeks in the simulation. Days of a week may have different activities but the same day in each week in the simulation has the same activity pattern unless the event of infecting occurs which changes the pattern of his/her activity.

As mentioned previously individual perform his/her activities engaged with a group. By changing the activity individual may join another group. This is considered to be one of the reasons that result in spreading the disease from a group of individuals to another. This type of changing activities called “between activities” interaction. Between two groups, an individual may be the connector. For example, the child student who belongs to both the family and the school class is linking the class group and the family group.

In addition to individual activities, the model contains group activities. The activity is considered to be “group activity” if at least two individuals follow the same activity pattern together. For instance, in the weekend, the household group will participate in the whole day activities together, engaging with other groups.

The activities along the model are of two types, fixed activities and non-fixed activities. Within weekdays, some of the activities are fixed such as going to school or work. On the other hand, the social activities are non-fixed and varying from a day to another. However, next weeks the same activities will exist on the same day.

3.2.2. Agents’ behaviour (Social Interaction Model)

Within activity (group) an individual may have full or partial social interaction with the other individuals (Bian 2004). Full interaction means that individual has contact with the entire group individuals (see Fig.3.6.a). An example of such type, the family group, which its members have contacts with all other individuals in the group. While the partial interaction means that, an individual with a group has contact with some of the group’s members but not all of them (see Fig.3.6.b). For instance, in a class group, a student may has contact with two or three friends and classmates not the whole class.

Figure 3.6 both full and partial Interaction: face represents an individual, circle around the individuals represents group boundary. Straight line between two faces refers the contact between them(Bian 2004).

An activity place has a couple of constraints. These constraints might be physical constraints like for instance the size, shape, or distributing the individuals (randomly or regularly), or they might be social constraints. The social constraints can be for example: the default social acceptable distance between two individuals, individuals can move and change place during performing an activity. The constraints play a basic rule in the individuals' social interactions. Moreover, these will determines the risk of the place under the term of spread diseases. For example if a rule of distributing of a place says that minimum distance between two individuals should be 2 meters, it is far enough to protect a susceptible from getting ill.

The social interactions among individuals occur while they are within an activity (a group). Individuals are constraining by the social and the physical constrains of a place of an activity while they are interacting. For example, during attending a lecture in a class, individuals student are less interaction with each other than break time. This because, they have been constrained by the distribution of them in sets, they are listing to the lecture, and most time silent. However, for some other classes such as art class they are more free to move and choose a place and have more interaction, and so on.

3.2.3. Disease model

The model consists of two parts: infection model and illness evolution. The infection model is running parallel to the interaction model. While the illness evolution model will start when individual, get the infection until recovered.

As been clarified in chapter two, Pertussis is a highly contagious disease. Patients are infectious during the early, catarrhal phase of the disease and remain infectious about 5 weeks. During this period, when an infected person cough, sneezes or even breathes, tiny droplets are released and can easily be inhaled by other people around.

While individuals performing their activities the disease model is running parallel with the social interaction model. An individual interacts with other group members either fully or partially. When two individuals come into contact, the disease model checks their infection and vaccination statuses. If one of them is infected and in his/her infectious period and the other is susceptible then a diffusion of the disease may occur.

In order to get an infection, an interaction between an infected individual and susceptible individuals should occur as a direct contact (no obstacles between them) after one week of catarrhal period starts. This contact is considered to have the following features that let the disease transmit from one to another:

- Distance: is less than or equal 1 meter.
- Contact type: is close direct.
-

Then the duration of contact will be computed. If it is more than or equal 1 hour then the transfer of the disease is executed. Otherwise, the frequency of contacts between these two individuals per day should be computed and the end of the day the duration of each of repeatedly contact will be added up. If the total time of all of meetings is equal to or more than one hour then the susceptible individual becomes infected otherwise no infection occurs.

The immunity level of individual determines whether the individual is susceptible and what the severity of disease will be is. This important factor is determined by the vaccination status of the individual. The following diagram (see Fig 3.7) shows the vaccination level and shows how each level leads to a specific degree of severity of the disease:

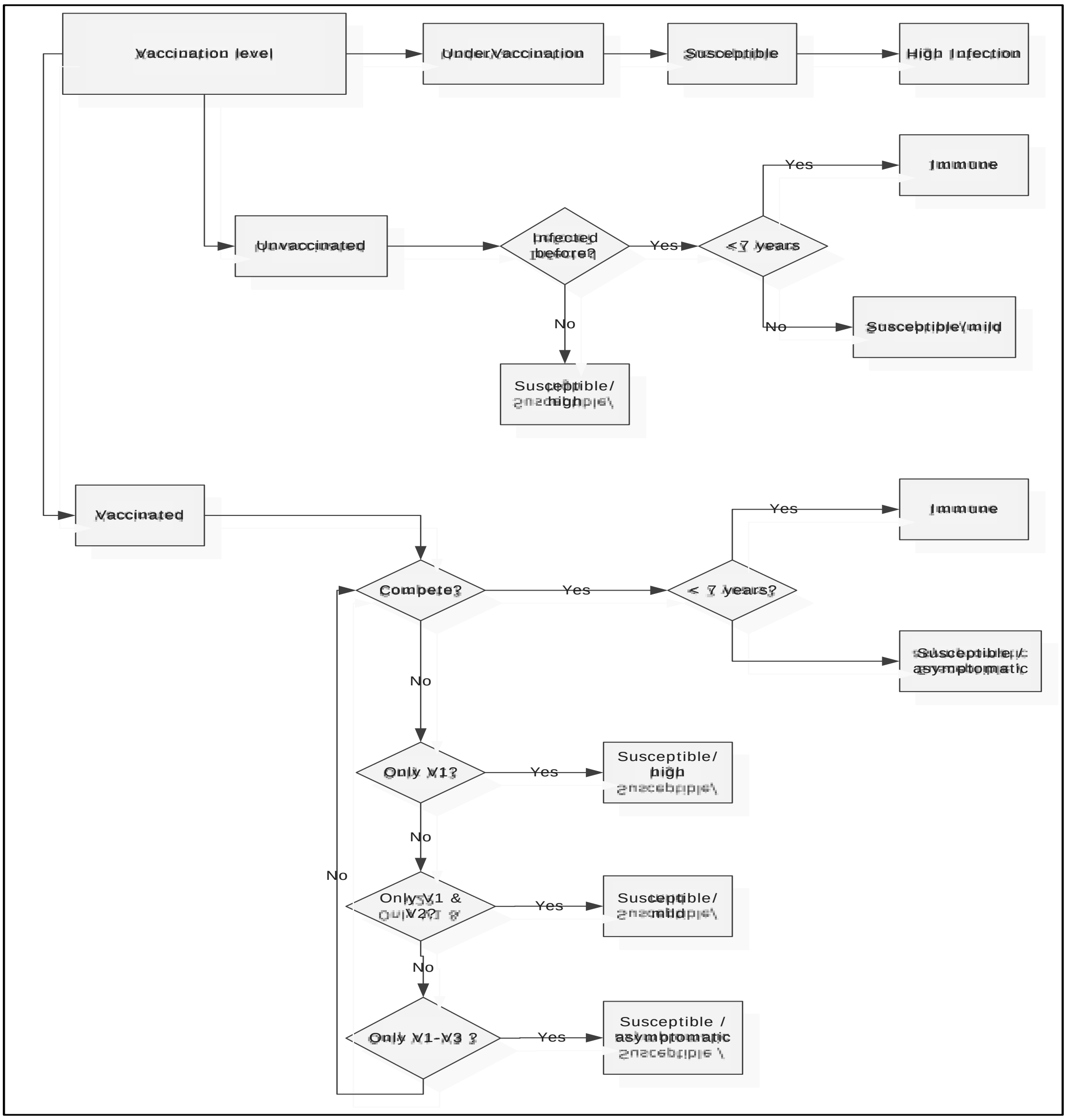


Figure 3.7: immunity, and infection levels based on vaccination level

This level of vaccination will be determined based on the estimations and proportions of surveys performed for The Netherlands. The vaccination level or status is one of the individual's attributes.

For declaring and simplifying, some assumptions will be made to cover some issues of simulating the disease spread:

- The contact is direct and includes face-to-face contact, physical contact (skin-to skin).
- For partial interaction, the individuals who are not in contact with the infected individual, if they shared the confined space in close proximity for a prolonged period of time (1 hour or more) they may get the disease.
- Infected individual: is the individual with disease in its infectious period of evolution.

- Child's families will consider immune after its child infected unless another child get disease from another resource.

The following diagram (Fig. 3.8) illustrated the disease transmission from an infected individual to another based on the latter immunity level.

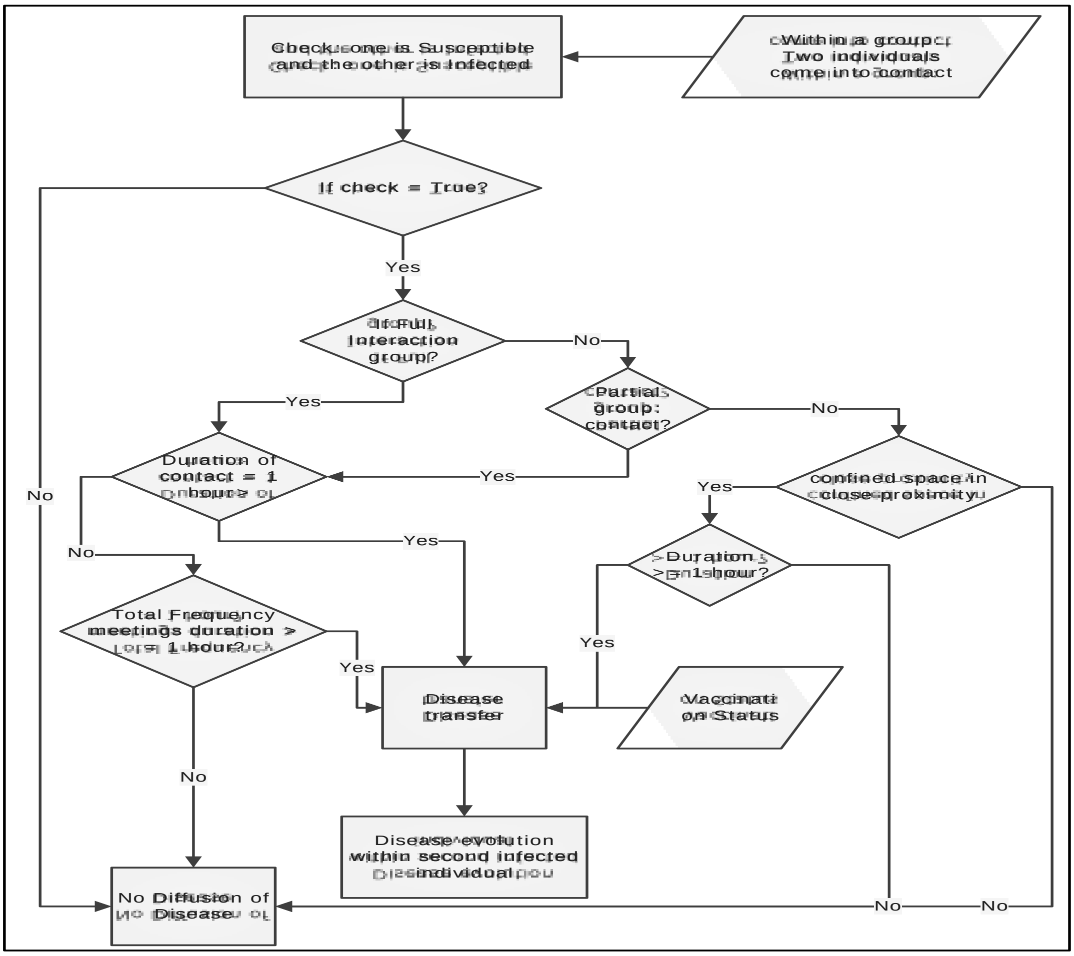


Figure 3.8: disease transfer between two contacted individuals

After susceptible individual becomes infected, the disease evolution passes through three stages of illness (see Fig.3.9). Before these stages, the incubation period (5 – 21 days) comes following exposure. The disease evolution of an infected individual starts and ends through these periods:

- Catarrhal: 1 – 2 weeks
- Paroxysmal: 1 – 6 weeks
- Convalescent: weeks – months

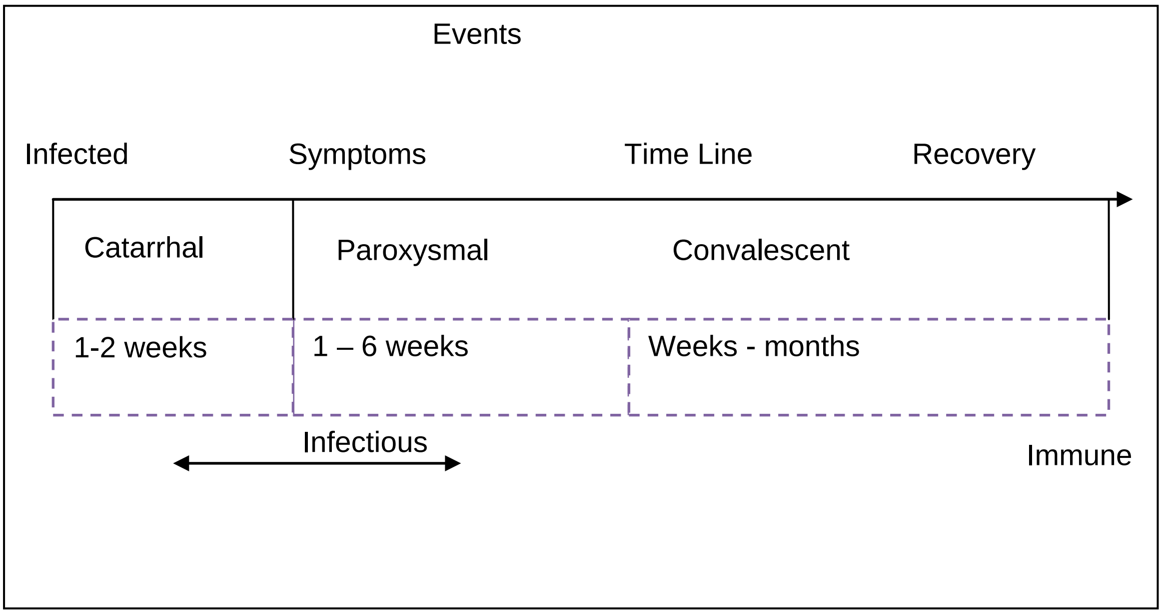


Figure 3.9 3.10 The Pertussis Illness stages

Infected individual in the incubation period will not spread the disease to others. Infectious period starts after one week of catarrhal period started and continues to 3 weeks after that. The moment the paroxysmal period starts within the event of occurring symptoms of Pertussis; the individual's attribute health status will be set to one. Then the sick individual will stop performing activities and stay at home. This will continue until recovered from the disease. After this event, the individual will be considered to be immune and will not be infected again. The level of illness and severity of the infection will depend on the age and again on the vaccination level.

4. Simulation data

Chapter 3 presented the conceptual design for the pertussis model. This chapter is discussing the preparation for the data that would be used for the simulation. Three types of data are used as model input: Enschede Census data, Enschede spatial data and pertussis vaccination data. The chapter was divided as following: section 4.1 discusses the data preparation of the Enschede census data. In section, 4.2 the preparation of Enschede spatial data is presented. Finally, section 4.3 discusses the pertussis vaccination data and the details dealing with this dataset.

4.1. Enschede Census Dataset

The first dataset is the census dataset for the Enschede region. The dataset was downloaded from the Enschede Municipality official website (Enschede-Municipality 2009). This census dataset was provided to the municipality by annual I&O Research. On a yearly basis, I&O Research[1] generates the statistical book “Enschede in Cijfers”. This dataset contains a summary of the local society such as ethnic groups and marriage status.

Each chapter contains a statistical aspect of Enschede population such as their geographical distribution over the 69 city neighbourhoods based on age groups, households, family sizes, types etc. In addition, there are chapters about city’s housing, economy, education and so on. The tables’ text information is in Dutch.

Data derived from this source are population per neighbourhood and family structure of the year 2009. These data were translated to English and then organized in “*.CSV” file format (comma-separated values). The reason behind choosing the data of 2009 is that this dataset includes detailed information needed for the model such as the number of families per each neighbourhood. The website of the municipality provides a page, which contains some statistical information so that a user can make a comparison between two periods. However even with the existence of some data of 1996, but it does not contain the mentioned details. They contain general information that helps in comparison and checking some information. It is good to state here that the talk about data of 1996 particularly because the empirical data used for the model are for the year 1996.

Comparing between 2009 and 1996 revealed the following: the total population of Enschede city in 1996 was 147,919 and it increased to 156,089 inhabitants in 2009. During these 10-years of period, 8,170 inhabitants increase the population. However, this increase is not equal for each neighbourhood. By looking at (Figure 4.1), the figure shows that between 1996 and 2009, the population of some neighbourhoods were increased and others did not.

---

[1] I & O Research is a multi-agency policy and market research, statistical analysis and advice see link http://www.ioresearch.nl/

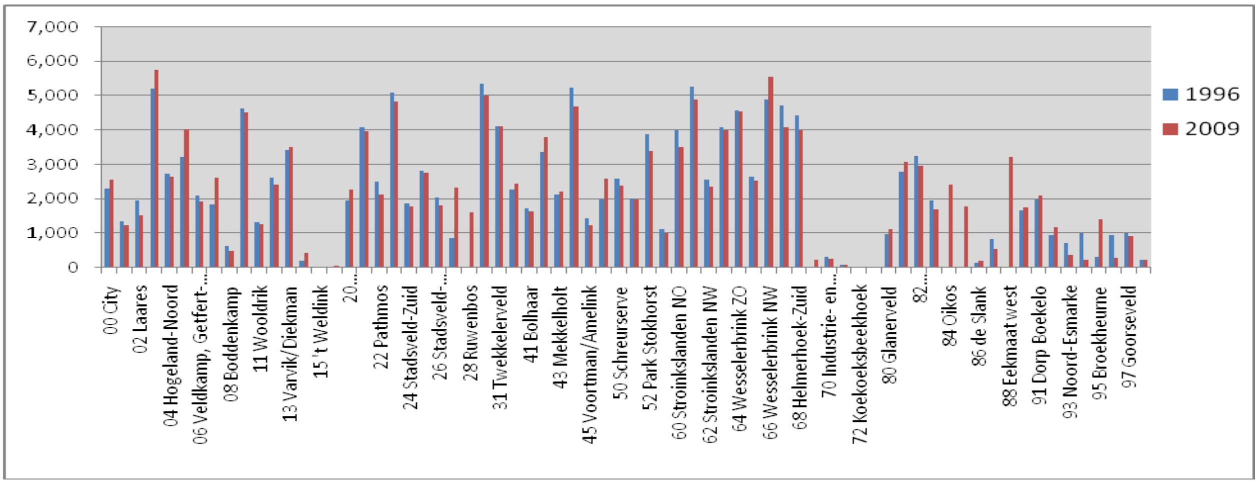


Figure 4.1Population of Enschede per neighbourhood in years 1996 and 2009

The total population of Enschede for the year 2009 was circa 156,089. Although eventually the simulation will have to run for this number of agents (or even far more) for the development of the system a smaller number of agents were used to avoiding the system delay during running the simulation. During the data preparation step a subset of data was generated from the downloaded population informed based on the purpose of the model, which is the simulation the pertussis diffusion among people of Enschede and exploring the patterns of this spreading.

The sub sets of data will represent the people of the city, which will apply their daily activities producing patterns of their activities. Tracking these patterns then change related parameters such as the vaccination level of the population would lead to explore the effect of these parameters on the spreading of the disease. Pertussis is a childhood disease and children are the risk group of infection. However, one of the sources of getting the disease is their family' members. It is infecting adults and adolescents. Moreover, the simulation system deals with smaller number of agents.

Due to avoiding delay during running the simulation and the model is childhood disease simulation, the population group, which has been chosen from the census data, is of families with children.

The families are of three types: family of couple of parents, family of single father and family of single mother all with children (see figure 4.2).

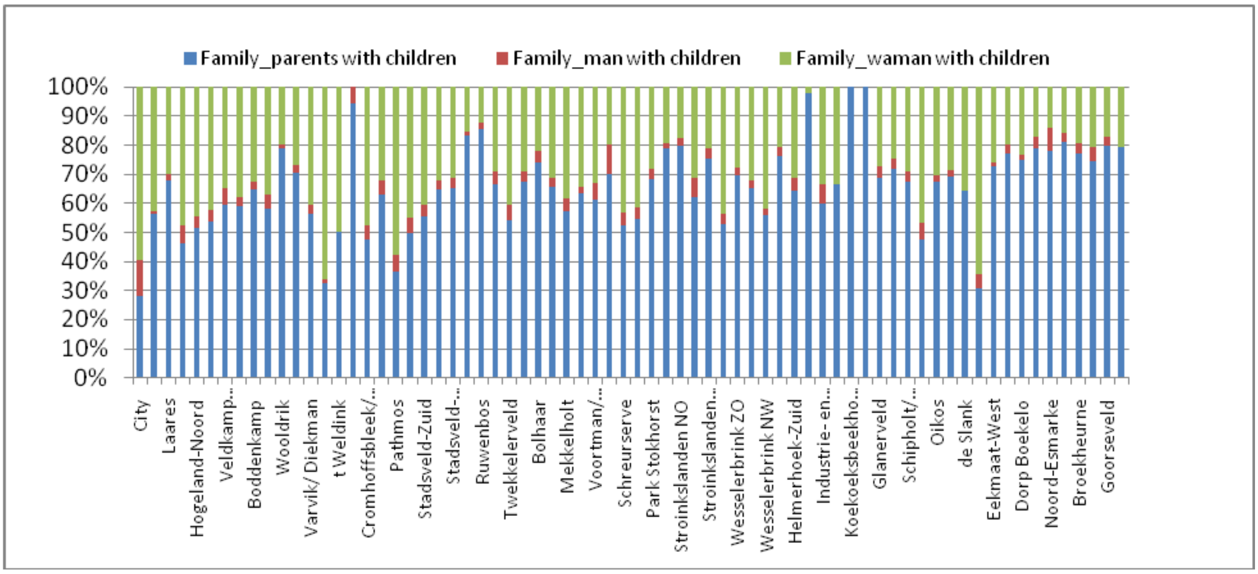


Figure 4.2 : Family percentage types Distribution over city neighbourhoods

The age groups are the four groups of children age groups: 0 – 4, 5 – 9, 10 – 14 and 15 – 19 years old (see figure 4.3). However, no further information on the ages of children is given that can help in distribute individuals age with the exact number. This means that for example, if age group (0 – 4) contains 100 individuals, we cannot know how many individuals are of age 2 years old. This is considered one of the limitations of the data.

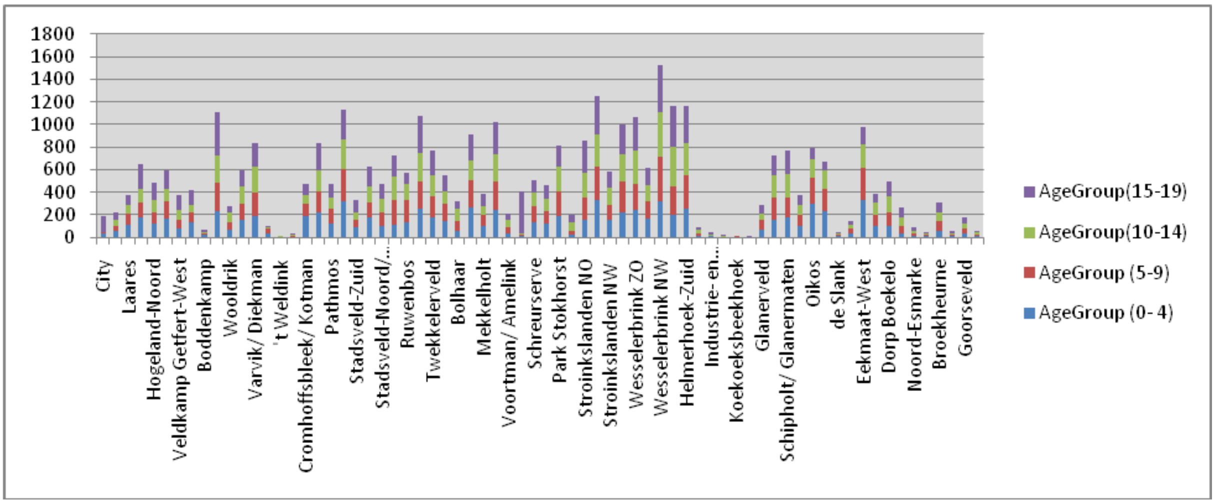


Figure 4.3 Distribution of Children age groups over neighbourhoods

The parents' age is assumed to be in the range (25 – 55) years. This adults age range is a good guess considering the children’s ages and it also matches the age range of the majority of adults having contracted pertussis base on the pertussis data presented by Van Boven, De Melker *et al* (see figure 4.4)

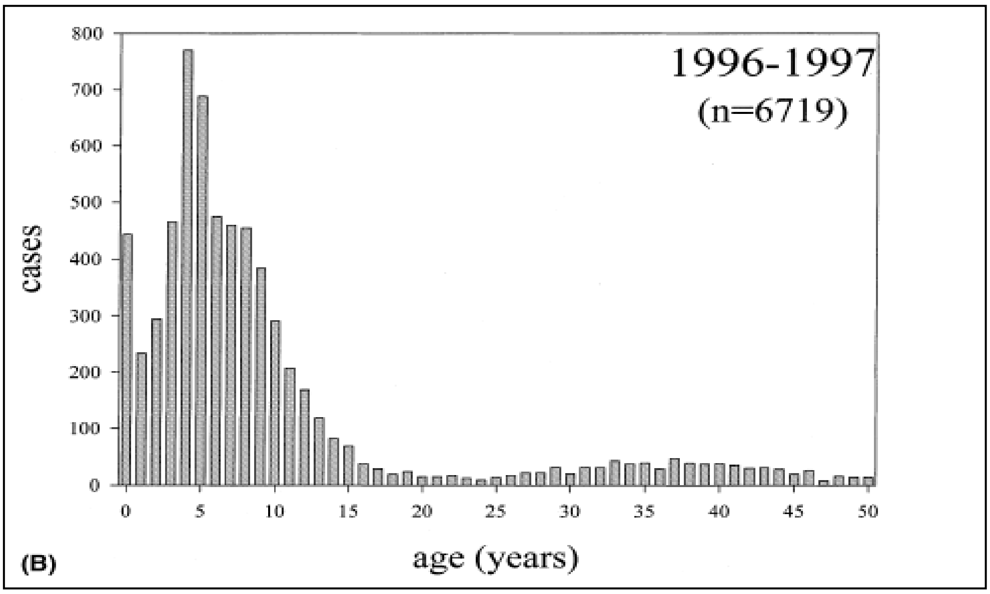


Figure 4.4 The observed number n of reported cases stratified by age in 1996 and 1997 (during the epidemic) (Van Boven, De Melker et al. 2000)

One of the type of demographic information provided within the census data of Enschede city is the economic data. These data provide the model with information of the number of employees in the city, both the full and part time work. Figure 4.5 shows us the total number of workers and the total number of full time workers and part time in Enschede in year 1996. 57.9 % of full time workers are men while the rest ratio is women. On the other hand, 39.2% of the part time workers are men and 69.8 % of them are women.

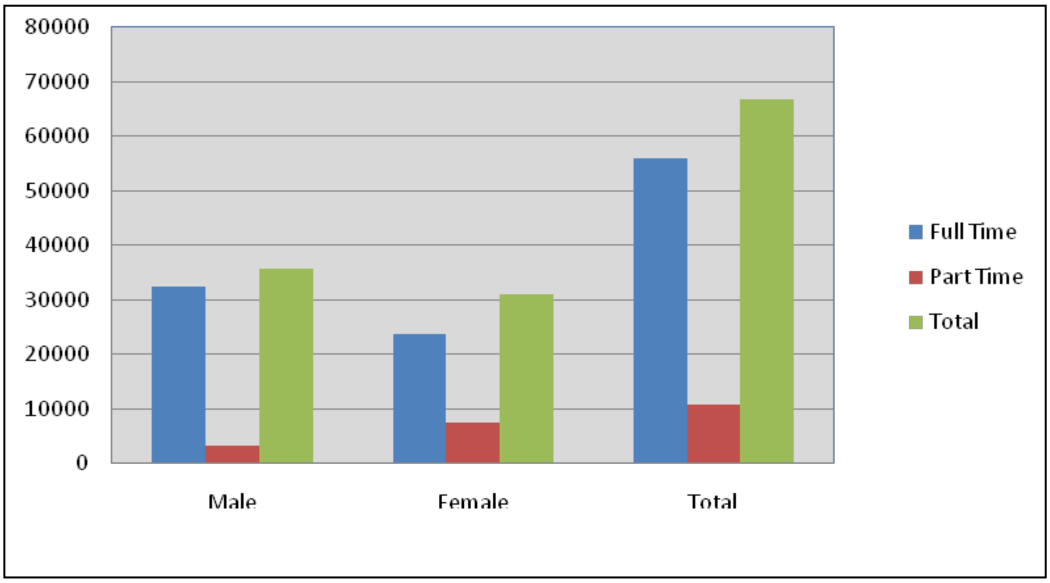


Figure 4.5 Workers in Enschede city for the Year 1996

During 1996 a total number of 66700 persons are working in the city weather these people are from the city or come from other cities is unknown. These people work in different places in the city with different type of jobs (see figure 4.6 which shows their geographical distribution over city neighbourhoods). However, the model named them workers and they work in workplace whatever this workplace is but it distinguishes among the workers of full time or part time jobs.

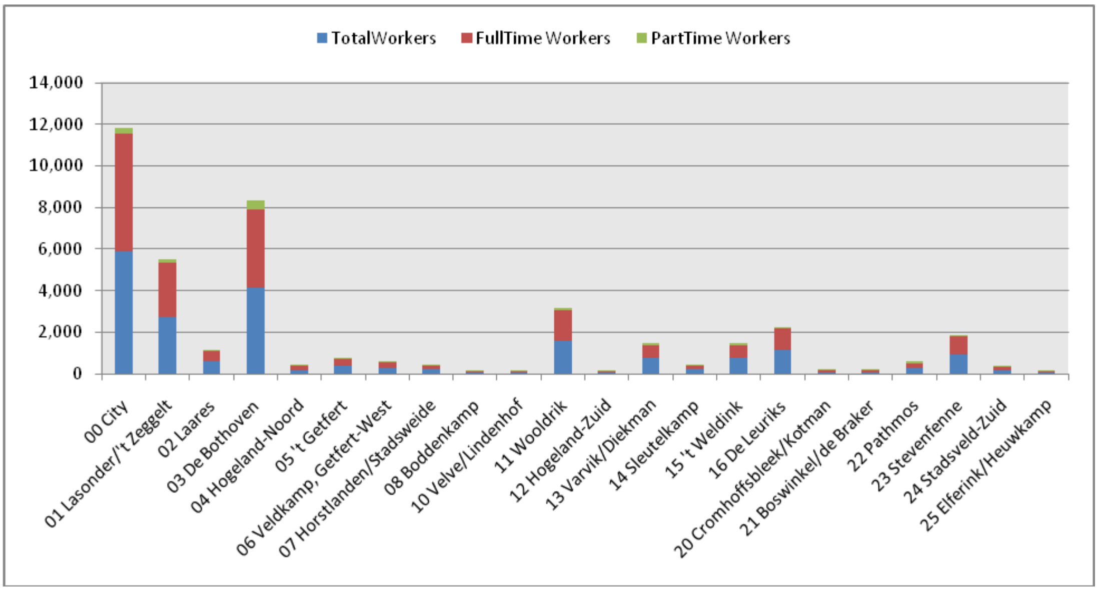


Figure 4.6 Distribution of both type of worker over city neighbourhoods

This distinguishing has a benefit of showing different patterns of activities during the simulation time. In addition, the economic level of the family whether both parents work or only one will reflect on the family's children activities. Both worker parents mean a family with high income. These two issues have more explanation in section 4.3.2.

4.2. Enschede Spatial Data

To let the agents interact with each other and apply their behaviour spatial data are needed to build an environment. The environment in our model is the Enschede city spatial structure. It should be known that the agents and the environment are strongly connected. For each purpose of modelling, agents are interested in specific objects of the real world. In our model, the interested object of the real world, which should be represented in the model, is the location (buildings). The locations (e.g. buildings) are a crucial part of the environment because they are the destination of individuals to meet up with other individuals who have the same activity. The spatial data that were needed for the model consists of two parts based on the purpose of using in the model:

- Spatial data used for building the model environment, this consists of :
  - City Neighbourhoods
  - Buildings:
    - Residential buildings
    - Schools ("kinderdagverblijven", elementary schools, high schools)
    - Work locations
    - Others (Shopping Centres, Sport Centres, Churches, etc.)
  - City business areas
- Spatial data used for visualization purpose:
  - City main roads
  - Water surfaces

The first group of spatial data were used to give the model agents the geographic addresses (Coordinates) of the places of their activity (houses, schools, shops etc.). The buildings’ feature used is the centroid point of them not there polygons. The second purposes of using spatial data is for visualization purposes in order for the user to get the idea of the city’s shape and features such as the main roads, and the buildings are included too.

The spatial data provided was by ITC. However, many important details were missing including building neighbourhood address and type. Let us explain each group with its content.

The *City_Neighborhood.shp* is used in the simulation to provide the model the city boundaries and the city neighbourhoods. This city boundary will help in determining the coordinates of the city that will help to simulate the proportion of the workers, who are working outside the city.

The *City_Buildings.shp* is used within the model to give the agents locations for applying their activities during the simulation. The *City_Buildings.shp* contains the polygons of the Enschede city buildings. These features have some attributes such as Id, shape, shape area, shape length, etc. however; the Shapefile does not have two main attributes "*Neighborhood_Address*" and "*Building_Type*". These two attributes could have been used during the implementation to assign the activity locations. The first attribute "*Neighborhood_Address*" is important be exist because based on the address of the buildings individuals could obtain house buildings for themselves and their families. On the other hand, "*Building_Type*" should exist to tell the system the function of the building to be used in the simulation. For example, the building type “School” could have been be assigned for the activity “going school” and it should not be used for residence purposes.

To determine the functions of use, they should be distinguished into residential and non-residential buildings. By assigning the buildings of (schools, high schools, work places etc.) the rest could be assigned as houses where the agents grouped within their family groups.

ArcGIS was used to assign the values of these two attributes. The buildings framed inside the neighbourhood's boundary would be selected then the neighbourhood name would be assigned as a value for the selected buildings' attribute "*Building_Address*". The name of the neighbourhood, which become a building type attribute values must match individual's address in order to use the building. This address will assure that create individual can find a house to live in.The building type is the key of identifying the building purpose of use. Therefore, the buildings types should be assigned too.

Determining or naming the non-residential buildings was done by using the telephone guideline website of the Netherlands (see: http://en.detelefoongids.nl/). By querying a required keyword of (What and Where) to the website search engine places would be returned as a result. Each of these places provided with address and a map. By comparing the website map to the ArcGIS map (consist of layers: boundaries, neighbourhoods and whole city buildings) then looking for the neighbourhood where the address zip code exist as well as the satellite layer map provided by the website would be used to give the real shape of the building. Because the normal map provided contains only the city roads and the landscape layers. The model map (the ArcGIS map) would be compared to the website normal map, to assign the area where the searched building located in. Then the comparison between the building shape in the model map and the existed one in the satellite map will help to find the building. The important places needed for the model are schools and work places.

For the first type of building, which is schools, it is important to differentiate between day care centre ("kinderdagverblijven"), elementary schools, and high schools (see figure 4.7). The importance of this differentiation is because different age groups attend these places.

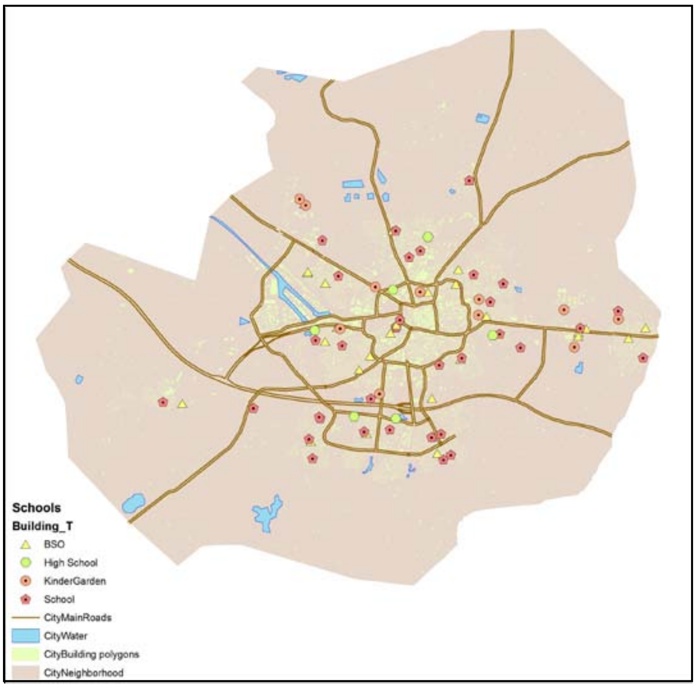


Figure 4.7 The geographical distribution of school types over the city

While for the workplaces, for example all industrial, commercial, and other offices, these are considered just as work places i.e. no differentiation among them (see Figure 4.8).

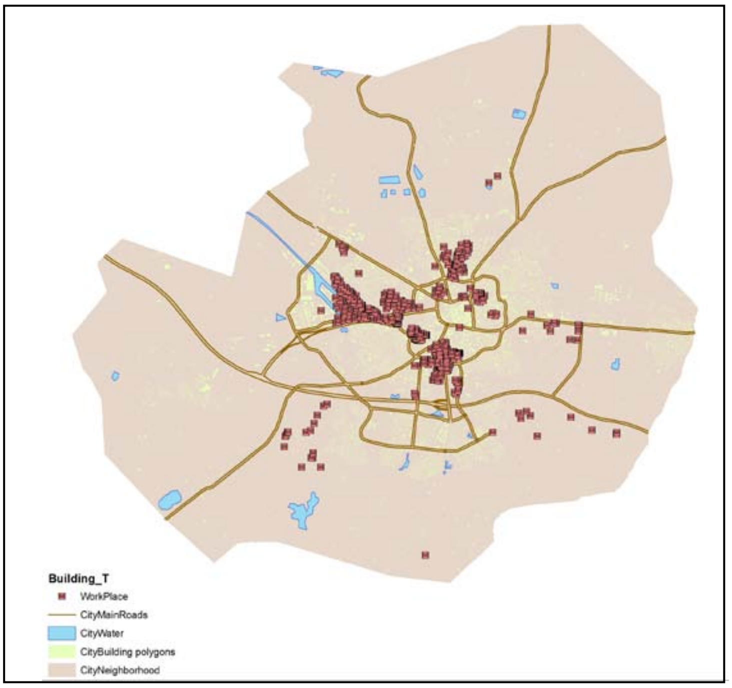


Figure 4.8 Geographical distribution of the workplaces over the city

In addition to finding these two important places, some other places were assigned such as sport centres, clubs etc. just by randomly assigning some points in different neighbourhoods as an alternative to tracking them on the phonebook. The individuals to practice their social activities use these places. For instance, a child may have a sports class once a week; in this case, s/he should apply that activity in one of these places. However, the type of these social places were distinguished and named according to the type of activity. For instance, the sports centre is given the name “sports” and is thus distinguished from “club” although both are used for social activities. Figure 4.9 provides the geographical distribution of these "social" places.

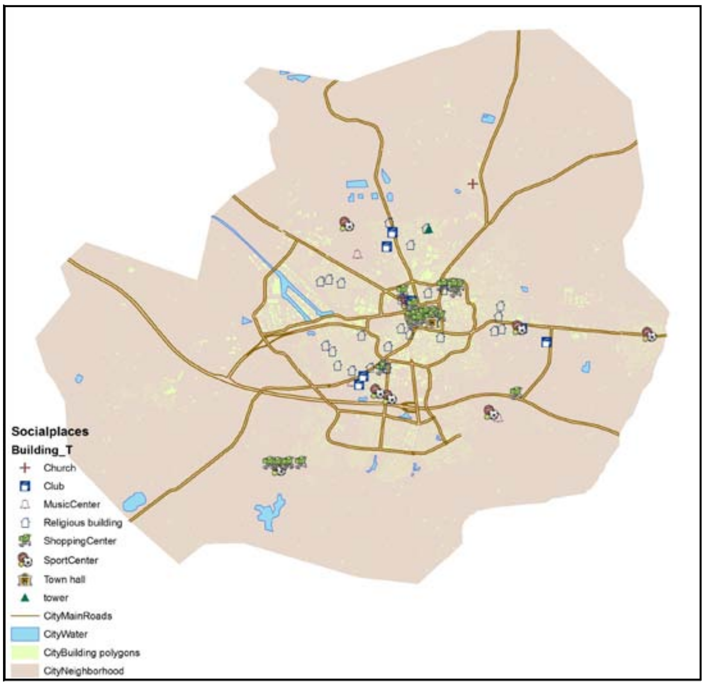


Figure 4.9 Geographical distribution of social places in the city

In addition, one of the data layers that have been provided was the business area layer. By overlapping this with the buildings layer, we labelled the “workplace” buildings, within the business areas “ shopping centre” some of the buildings situated in the city centre.

The Shapefile, which consists of the polygons of the city buildings was used for visualization during the displaying of the simulation. However, this Shapefile was the original of the one which is used in the implementation. The centroid points of these building were derived from this Shapefile. The model uses the centroid points in its computation. These points were extracted using ArcGIS tool function (Feature *to Point* method). Later these points will be loaded to RepastS.

It should be stated that although the spatial data had the information about the area of its shape but in this step of developing the current model the building’s area or departments did not be considered in the calculations of the model. Hence, more data that are detailed required which currently not available to get use of them in the distribution of the agents groups within one building. For instance within an elementary school of at least 8 different classes, all of the classes groups assigned to the centroid point of the school without any spatial distance of obstacles split them into groups or classes. The area of the building shape that is provided is not good choice to divide a building to departments. That is because a building may have a small area but it includes couple of floors, which are splitter between different groups sharing the same building.

The second group of the spatial data set was used for visualization purposes during displaying the simulation. These data did not have any effect on the individual agents. They were added at runtime of the simulation. In addition the simulation has the city main roads Shapefile data but the individuals do not use this spatial information to move from location to location. The way the movement behaviour is modelled is that they just disappear from the current location and appear at the destination location. These groups of data did not need any preparation.

4.3. Pertussis and Vaccination Data

The model is simulating the disease spread. This disease spread from an individual to another is based on contacts (existing in the same place, same time and same group), infectious level of the infected individual and the immunity level of the receivers. It is important that each agent in the model have a level of susceptibility. This level determines how likely it is that this agent will be infected when in direct contact with another infected agent.

The agent is considered to be susceptible if s/he is not vaccinated against pertussis, has un-completed vaccine or waning of vaccination. Even if the person had been vaccinated, the vaccine might not give the person the required immunity -and that might happen- because of the effectiveness of the vaccine itself. This explanation is according to the reported and serologic data and the hospital admissions, which said that the increment in pertussis occurrence was higher among vaccinated person than non vaccinated person of all ages (Hester E. de Melker, J.F.P. Schellekens *et al.* 2000). In effect, vaccination never means that the person is completely protected from getting and spreading the disease.

Estimates for induced protection is of limited duration in the range of 5-10 (or 6 to 8) years and individual efficaciousness of pertussis vaccines are around 70-80% (Taranger, Trollfors *et al.* 2001; Wirsing von König CH, S Halperin *et al.* 2002; Netherlands 2004; Águas, Gonçalves *et al.* 2006). In addition, waning vaccine-induced immunity has been suggested as another explanation for the occurrence of the disease as an epidemic in vaccinated population (H. E. de Melker, M. A. Conyn-van Spaendonck *et al.* 1997; Águas, Gonçalves *et al.* 2006). The waning vaccine-induced immunity led to pertussis shift to older age groups in addition to the infants (Christie, Marx *et al.* 1994; Skowronski, De Serres *et al.* 2002).

It should be declared that the natural infection and vaccination induce immunity are assumed to give similar immune protection and assumed to decline at similar rates (Cherry 1999). This means that we might consider an individual previously infected with pertussis, s/he would consider being immune for 5 to 10 years then loses immunity and then individual again became susceptible then may be reinfected. This reinfection of individuals with maximum levels of protection can potentially appear and result in mild or atypical symptoms (Christie, Marx *et al.* 1994; Wirsing von König CH, S Halperin *et al.* 2002). Crowcroft and Pebody (2006) argued that immunity has a great effect on severity of the disease and a limited effect on transmission of infection. Even though they have mild or atypical symptoms pertussis, they have the capability of transmitting the disease to other susceptible individuals especially infants within their households(Águas, Gonçalves *et al.* 2006). As a conclusion the fact that we retain for our model is that immunity to pertussis after both vaccination and natural infection wanes after an unknown number of years.

Vaccination data, both for Europe and for the Netherlands,was extracted via literature (De Melker, Spaendonck *et al.* 1997; Boven, Hester E. de Melker *et al.* 2000; HCN 2004; Lee, LeBaron *et al.* 2005). This information was used to determine the proportion of the population that we would consider to be vaccinated and their level of vaccination. For instance, it was considered that 96% of children of age 4 have their complete vaccination (HCN 2004).

The historical data about pertussis cases having occurred in the Netherlands, provided by ITC contains information about where (zip-code) and when (date) the case has been registered in the epidemic years from 1993 to 2004. These data are available for the Enschede region. However, it is incomplete in the aspect that does not provide the age profile of the patients. The age profile of the recorded cases is of high importance. It is closely related to the transmission of the disease (Águas, Gonçalves *et al.* 2006). In addition to the absence of age profile within the empirical data, no information about the severity of the cases is provided. The data is used to verify the disease model to see how close the model is to reality. That offer a requirement to extract data from literature, which showed where the infection density per age is concentrated. Also peaks of epidemics where were identified.

Based on vaccination data of the past 18 years, for all children, an immunity level can be calculated making the following assumptions:

- Infants of age younger than 6 months are considered as under vaccination age, however; within the new type of vaccine they try to include infants of 2 months of age.
- Children under the age of four but older than 6 months (96% is vaccinated and all vaccination is partial as the full booster has not yet been applied).
- Children between 4 to 9 years old (96% is fully vaccinated and the immunity has not deteriorated because of waning) this is the best-protected group.
- Children between 10 -18 years old were vaccinated but with old serum (not the new booster) and their immunity has waned by unknown level.
- Adults who are older than 18 years old assumed to wane their immunity against pertussis. We assume that 10 to 30 % of them have had pertussis (Hoey 2003).

Adjustment for the immunity caused by natural infection, pertussis data are available for the past years. However, the following comments have to be made:

- Only cases that are diagnosed as pertussis are included in the historical dataset. When people are (partially) vaccinated but their immunity has waned they will get a milder form of the disease that might not be recognizable as patient with pertussis.
- Obviously, the data contains only patients who visited a hospital or a doctor and have been recorded there. In addition, not all general practitioners have all possible patients tested. This means that the number of pertussis cases recorded for the Enschede region are inferior to the total number of pertussis occurrences .
- The data did not provide the patients' age. There is no information whether the patient was infant, child or adult.

In the simulation the immunity level of agents depends on their age, the data from literature, and the census data. Besides, in order to address the effect of immunity level of the population on the diffusion of the disease the model could have a parameter called “vaccination” increasing or decreasing the ratio of the population immunity. During runtime of the simulation, this parameter's value could be changed and analyzed of each run to show the effect of the value of the parameter in the simulation. How these data have been employed in the simulation was provided with more explanation in the following chapter

5. Implementation of the disease model

In chapter 3 the model conceptual design was presented. The model data description and preparation was presented in chapter four. The next step would be the implementation of the model. The model was implemented using the agent-based simulation toolkit Repast Simphony and Java Topology Suit under the object-oriented programming style of coding using pure Java programming language. The chapter is divided into sections and subsections. Section 5.1. presented an overview of the modelling tools and concepts related to Repast Simphony that are essential for the implementation. In section 5.2, the actual implementation of the model related to the simulation stages of the conceptual designed model is presented.

5.1. Tools for Implementing the Model

The agent-based simulation toolkit Repast Simphony (or Repast S) and Java Topology Suite, which is one of Java API, are the basic tools of the simulation. Java programming language is the objected-oriented language of the simulation. The following sections will provide a brief overview on these two tools.

5.1.1. Repast Simphony Software Framework

Agent-based modelling has numerous toolkits such as Swarm, NetLogo, MASON, OBEUS, and Repast, which support the direct integration of geospatial data. An evaluation of freely implementation libraries Swarm (Java version), Repast and two other less-used software frameworks (Quicksilver, and VSEit) had been done by Tobias and Hofmann (2004) to determine the simulation framework that is the best suited for theory and data based modelling of social interaction. The authors concluded with the result showed that, the Repast environment is the clear winner under the terms of official program documentation, statements by developers and users, and the experiences and impressions of the evaluators (Tobias and Hofmann 2004). Two years later, Rails back *et al.* (2006) presented another evaluation. This evaluation which compared NetLogo, MASON, Repast, and Swarm (both Java and Object C versions) frameworks showed that Repast was considered to be “the most complete Java platform” and “its execution speed to be good compared to the other platforms” (Railsback, Lytinen *et al.* 2006).

The University of Chicago’s Social Science Research developed Repast or Recursive Porous Agent Simulation Toolkit in 2003. It is a free open source software platform developed as a pure Java implementation (North, Nicholson *et al.* 2006). Repast can be used for creating agent-based simulations and has been used extensively for modelling social behaviour however, it is used for other simulation purposes also (Macal and North 2005; North, Nicholson *et al.* 2006). The Repast Organization for Architecture and Design (ROAD) is responsible for development of the software (North, Nicholson *et al.* 2006). Three Repast platforms are available currently using three different implementation languages: Repast Py (using Python programming language), Repast Net (using any programming language compatible with Microsoft.Net such as Visual Basic.Net, C++, C#, J#, etc) and Java Repast (both RepastJ and Repast S using pure Java programming language). However, Repast Py, Repast.Net, and RepastJ are no longer being developed as they reached maturity. By the existence of

Repast Simphony (or the abbreviation RepastS) and its supply of all the core functionality of Repast.Net and RepastJ the mentioned platforms have been superseded (Crooks 2007).

RepastS provides functionality libraries of classes that create, run, display and collect data from agent-based simulations (Crooks 2007). Moreover, RepastS comprises such functionality as scheduling events, charting, record data, snapshots of a display etc. The availability of basic functions in the Repast library allow programmers to spend more time developing specifics of their models like agents behavior, interactions and etc. (Crooks 2007). RepastS supplies a more point-and-click interface of a certain part of the model. In 2006, the Beta version of RepastS was released but little details about all its core functionality are there. However, on the website http://repast.sourceforge.net/ the free software and the documentation of How-to are available and downloadable. The developer team of Repast has provided a couple of useful articles about the software and the functionalities it has. North *et al.* (North, Howe *et al.* 2005a) introduced the architecture and core functionality. While in (North, Howe *et al.* 2005b) the authors outlined features for data analysis (integration of the R statistical package) as well as the agents' interaction/behaviour, storage and display and presentation of models with RepastS too. Tatara *et al.* (Tatara, North *et al.* 2006) highlighted the ability of RepastS to create and display 2 and 3D by providing a detailed discussion showing how-to develop a "simple wolf-sheep predation" model. The development concepts of RepastS environment has been discussed by Howe *et al.* (Howe, Collier *et al.* 2006).

The implementation language Java of this platform adds more advantages to the software. Java is an independent platform "write once, run anywhere", which gives the advantage of running simulation on a computer with different operating systems for instance Windows™, Mac™, Linux™, Unix™, etc. The richness of Java over other programming languages such for example Visual Basic.Net is in its much larger set of third-party libraries gives Java faster development since in many cases the existed code can be reused (North, Nicholson *et al.* 2006). For the advantages of RepastS and its implementation language Java over other platforms, it has been used for implementing this simulation model.

5.1.2. Spatial Data Operation Tool: Java Topology Suite

For topologic calculations, the user has to link the ABM model to the Java Topology Suite (JTS). This linking is because RepastS does not support a means for the model to interact with Application Programming Interface of the GIS (Crooks 2007).

Java Topology Suite is a Java API that implements a core set of spatial data operations using robust geometric algorithms and an explicit precision model (Aquino 2003). A complete model for specifying 2D linear geometry is provided within this Java application programming interface. Many common computational geometry and spatial data processing operations are exposed in a consistent, clear and integrated API. For the applications which support the cleaning, validation, integration and querying of spatial datasets JTS is intended to be used. The main tasks of JTS involve creating and using geometry objects. By creating the geometries, many computations can be done such as intersection, overlay, area calculation, envelope, centroid etc. JTS also provide a set of Boolean operations which compute directly common spatial relationships (e.g. equals, within, contain, disjoint, touches, etc), overlay operations (e.g. buffer, union, intersection, etc) and so on. During coding the conceptual design of the model, the usefulness of this Java topology tool will be much clearer as in every step a use of spatial data exists. JTS will provide these basic common operations to handle them properly.

5.1.3. Concepts from Repast Simphony

Context in RepastS is the core data structure. Simply it is a bucket or container of all of the model elements based on set semantics representing an abstract population. To define a population in a model and the interaction of that population without actually providing the implementations, the *Context* provides the basic infrastructure. Objects (or agents, see 2.1.2) in a *Context* are model population. *Contexts* create an abstract environment where the agents at a given point exist in the simulation. *Contexts* could be city, buildings or even individuals. Simple values such as time can be handled within the *Context*, which gives the agents information about the world where they interact. However, any mechanisms for interaction between agents or concepts of space are not provided by the *Context*. There is no specification of relationships between *Context*'s members (Howe, Collier *et al.* 2006). *Context* can contain *sub-Contexts* if it is necessary. In a *Context*, membership is inherited. So if an object is in a *sub-Context*, by definition, it is a member of the parent *Context*. This hierarchical feature of the *Context* allows the designer to define consciously the granularity of the model. For example, the *PopulationContext* may have the sub-context *IndividualContext* that may have a *sub-context* of *FamilyContext*. Here the *PopulationContext* is the root *Context* of the *sub-Context IndividualContext* that is the father of the child *sub-context FamilyContext.*

As mentioned before, *Context* does not provide agents with concepts of relationships or space. After having agents in a *Context*, a space is given to them through a data structure called *Projection*.

Projections defines and applies relationships between a given Context's agents (Howe, Collier *et al.* 2006). A Context's *Projection* will automatically contain all the *Context* agents (Malleson 2008).
Projections are the keys to agents' interactions. They are added to a *Context* to let agents interact with each other. For instance an (x, y) spatial location can be given to each agent with GIS *Projection*. The relationships between agents (e.g. Social network) can be defined with *Network Projection*. A *Context* can have more than one *Projection*. This feature allow agents of a *Context* to have a number of types of relationships with each other (Howe, Collier *et al.* 2006). Switching between *Projections* of one *Context* is simple. It does not require any change with respect to agents' code for a *Projection* to be able to work. However for the agents' to work with a new *Projection* changes are required (Howe, Collier *et al.* 2006). RepastS has four types of projections each for different purpose of use: *continuous space projection, GIS projections, grid projections* used for Cellular Automata *and network projections*. A *Context* can have more than one projection that allows the context's agents to connect to them and apply their behaviours.

The spatial data (presented in section 4.2.1.) loaded to contexts geographic projections to provide the projection with the required information and data to build the environment where the individuals will use to move on and interact with other individuals. Later the network projection will be used for visualizing the social network of infected individuals.
It is important to say that in order to manage the model agents and projections, the "model.score" is the key providing Repast Runtime this ability (Liebert) which reads this file and knows what is there in the model. The model.score is an xml file describing the model components. In model.score file, agents, ***Projections*** can be created and modified for each context. It must be said that the names of ***Contexts***, Agents, and ***Projections*** in this file must match the associated classes. Just by creating the model project model.score will open automatically.

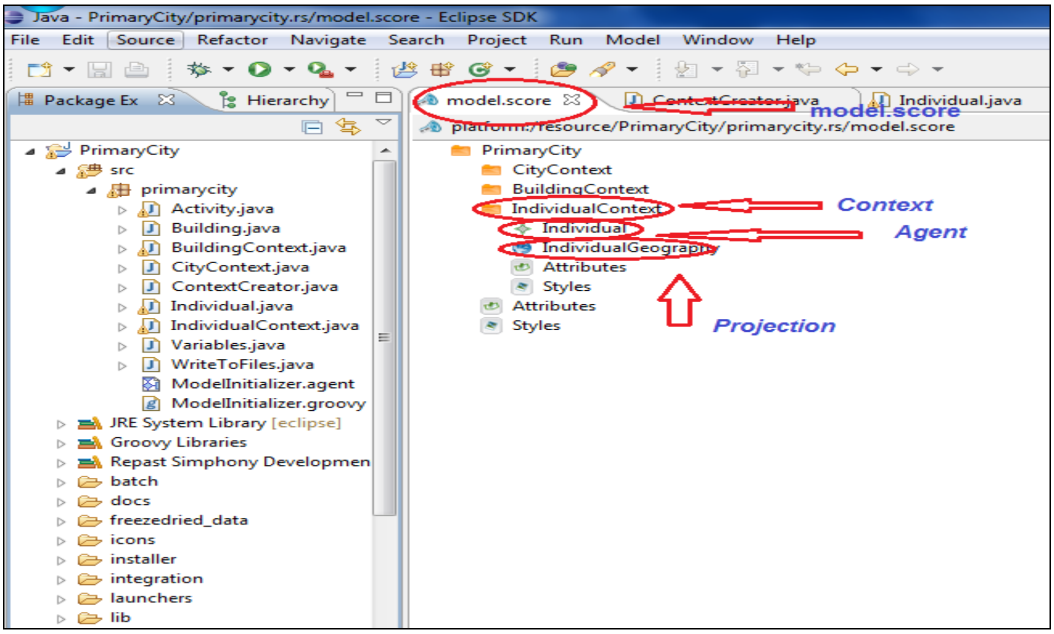


Figure 5.1 model.score contains (Context, Projection and Agent)

The execution of agents' actions (e.g. agents' behavior) and the model's actions such as display updating is done via the ***Schedule***. Repast present a mechanism for scheduling events that allows for more sophisticated dynamic *schedule*s to be coded such that the execution of an event can schedule itself over events for the execution in the future or in other words it provides event driven processes. However, the schedule is not responsible of all communication between model's parts. Scheduling in RepastS composes of setting up and executing actions at some specific time related to other actions, thus within ABMs this represents dynamics of the model. The method of the schedule during the simulation would be called at every time step. The time quantum unit in Repast is known as 'tick'. The tick is used as a hook on which the execution of events can be suspended, ordering the execution of events (Collier 2002). Ticks are only a way to order the execution of events related to each other. Repast makes scheduling explicit. The mechanism of scheduling of Repast is responsible for all the changes in user-defined state with a Repast simulation. It is a fully synchronous discrete event scheduler. The scheduler support both sequential and parallel operations of discrete events. According to Brown, Riolo *et al* (2004) implementing scheduled events may be implemented in three ways within Repast

1. Sequencing events in a synchronous step-wise fashion. For instance, each agent or a group of agents is marked to perform its task once at each time step or once every n time steps (ticks).
2. An event is scheduled to occur at a time step n only once. Within this fashion any number of different events that may be scheduled to occur providing an in advance history of events to take place.
3. Encapsulating 'event-driven by the model therefore agents trigger events to occur or add events to the queue of events (schedule) to take place (Ropella, Railsback *et al.* 2008).

The tick count is incremented when the schedule has finished its iterating over the execution queue. Within the use of this tick count, actions are scheduled for executing. "The ability to easily schedule events is of great value, not only to actions within the model such as updating the display, recording information and so forth but also to agents' states" (Crooks 2007). It should be said that Repast does not specify any mechanism for interaction between the model environment and agents that make up that state. No set formula for creating a model is provided by Repast: the modeller/programmer is specifying all required changes. It is the matter of deciding what is needed and relating to the appropriate classes (Crooks 2007). Therefore, this requires much effort to understand how to design and implement the model.

The importance of scheduler functions has come from the desire of knowing the circumstances or local interactions among agents that gives rise to a specific global structure (Sansores and Pavón 2005). The consequences of scheduling will help to have a control over the simulation to observe in every moment what is happening and which parameters influence the simulation results. In the case of our model, the schedule importance is for using schedule to implement the daily activities of the individual. Each individual in the simulation has a timetable including the time of moving or doing something at that moment. In addition, the scheduler will provide the system information about how long an individual has been infected then until what period of time individual should be present then stop activity and stay at home. Moreover, the scheduler will provide the individual with the information of "now two weeks of recovery is passed, you can go back to your normal activities". In addition, the schedule will help to distinguish among the weekdays and weekend days since the activities of weekdays are more individuals while in weekend the activities are group activities.

In RepastS or Repast Simphony, there are three ways to work with the scheduler. Each of them has its specified purpose and no one is better than the other (http://repast.sourceforge.net/docs/reference/SIM/)

- *Directly Schedule an action*: with a slight twist, this scheduling way is similar to the way that actions (events) have always been scheduled in Repast. Here in this way, the modeller get a schedule and tell it the "what when" should run. The slight twist provided here is that rather than using one of the numerous methods to set up the parameters for the event the modeller only use an instance of *ScheduleParameters*.

```
ScheduleParameters startParams =
ScheduleParameters.createOneTime(1);

schedule.schedule(startParams, this, "start");
```

Figure 5.2: an example of Directly Scheduling an action

In this piece of code (is a part of ContextCreator. see the Appendix) `"start"` is the schedule method that will be called one time in the simulation for the class context it has been written in.

- *Schedule with Annotations*: one of the most useful features Java 5 introduced is Annotation support. In Java, Annotations are bits of metadata that can be attached to classes, methods, or fields that are available during run time to give the system more information. Within RepastS, its developers thought that annotations were a perfect match for tying certain types of scheduling information to the methods that should be scheduled. The typical case where to use annotations is where there are actions whose schedule is know at compile time. So for example, if in our model we want to create a specific plan for individuals the annotation would be logical to be used. The following java statements declare how to schedule an action using annotation:

```
  @ScheduledMethod(start=1 , interval=1)
  public void start(){
  …
  }
Schedule.schedule(Individual);
```

Figure 5.3 Schedule with Annotations example

The properties for the annotations arguments and for the *ScheduleParameters* object are similar. Using annotations for scheduling has one particularly nice feature that it is easy to keep track of "what executed when" by getting keep the schedule information right next to the method that is scheduled. Objects with annotations will be added to the schedule automatically most of the time. However, it may not be the case when creating a new object when the simulation is running. Scheduling objects with the annotations can be made easy by using scheduled object. The last line in the above code the schedule will search the objects for any methods, which have annotations and add those methods to the schedule. It has to be stated that this type of scheduling is designed to handle only scheduling where the actions are well defined at compile time, but not the dynamic scheduling.

- *Schedule with Watcher*: the most radical of the new scheduling approach is scheduling using watchers. Watchers are designed to be used for dynamic scheduling where the model designer is well understood a typical workflow. Mainly, a watcher allows an agent to be informed about a state change in another agent and schedule an event to occur as a result. For example, in case an epidemic occurs, the agent government should be informed to apply some specific procedures, or might be parents of children will prevent their children to go to school until the epidemic is over. Like the java code above, the watcher is set up using an annotation. However, instead of using static times for the schedule parameters, a query is specified by the user which define whom to watch and another query which defines a trigger condition that must be met to execute the event. For better declaration, if we take a look at the following example:

```
@Watch(watcheeClassName = "repast.City.Individual",
watcheeFieldName = "illness", query = "getChild",

               whenToTrigger =
(WatcherTriggerSchedule.LATER, scheduleTriggerDelta = 1,
scheduleTriggerPriority = 0)

public void childsick(Individual child) {

    if (this.child.isIll())

        this.setbackhome(this);

    else

     this.continuework();

     System.out.println("checking children health
     status");

}
```

Figure 5.4 Schedule with Watcher example

Let us add some words for better explaining the code:

Watcher queries are Boolean expressions which evaluate the "watcher" and the watchee".

`@Watch` annotation, this tells the RepastS system that this is going to be watching other objects in order to schedule the "childsick()" method.

- The first parameter of the annotation is the `watcheeClassName` which defines the type of agents that this object will be watching.
- The second parameter, `watcheeFieldName`, defines what field the user is interested in monitoring. This need a variable in the class "Individual" which is wanted to monitor the changes. Whenever this variable value changes, the object will be notified.
- The `query` argument defines which instances of Individual we want to monitor. Here in this example, the system is monitoring agents whom are child of the agent "parent".
- `whenToTrigger` this argument specifies when to execute the action, whether immediately before the other actions are executed at the same time or to wait until the next tick.
- `scheduleTriggerDelta` defines how long to wait (in ticks) before scheduling the action.
- The last argument is the `scheduleTriggerPriority`. This argument allows the user to define the order in which this action is executed at its scheduled time.

After the above explanation of the Repast main concepts, the rest of the chapter was about the real implementation side of the pertussis disease model.

5.2. Implementing the three Simulation stages

The preparation of data was discussed in chapter 4. The tools of the simulation were presented in the beginning of this chapter. The next step is to implement the conceptual design of the pertussis model. The stages are creating synthetic population and loading the city environment, allocating activity tables to individuals and joining groups within an activity, and visualizing the social networks. The following sub-sections are going to describing the steps of applying these stages. For the time being, the first two stages (creating the model base line: agents and environment and enabling agents to behave and interact) are going to be implemented. The last stage, which is the visualization of the infected individuals social network might be the next future step.

The concepts presented early in this chapter (Context, projection and Schedule) will be used in the description of the model implementation to provide clear perspective on the way of implementing the model. Figure 5.5 illustrate the model implementation elements of the model:

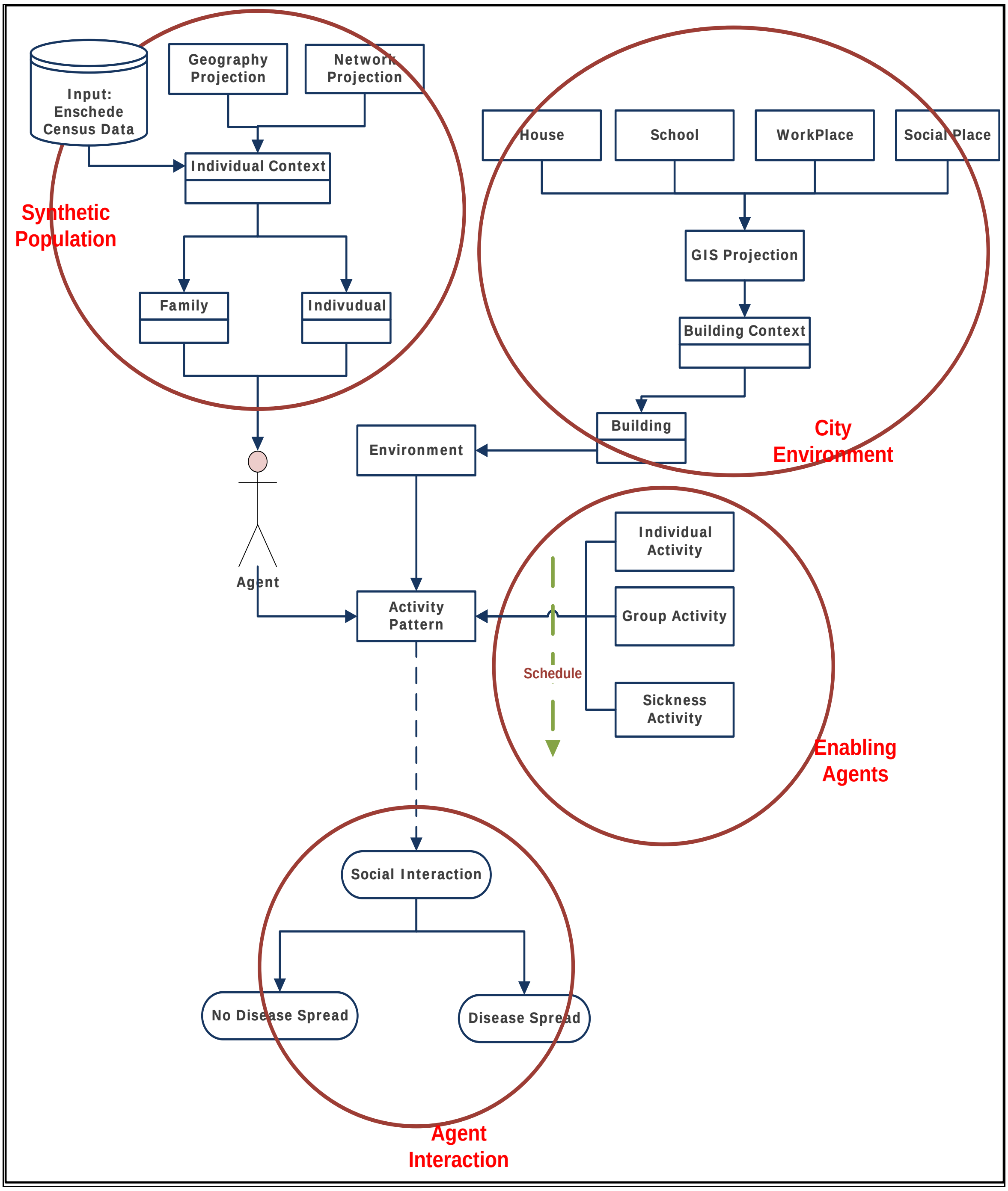


Figure 5.5 Model Implementation Elements

In addition, the following figure illustrate the UML diagram of the main classes in the pertussis model

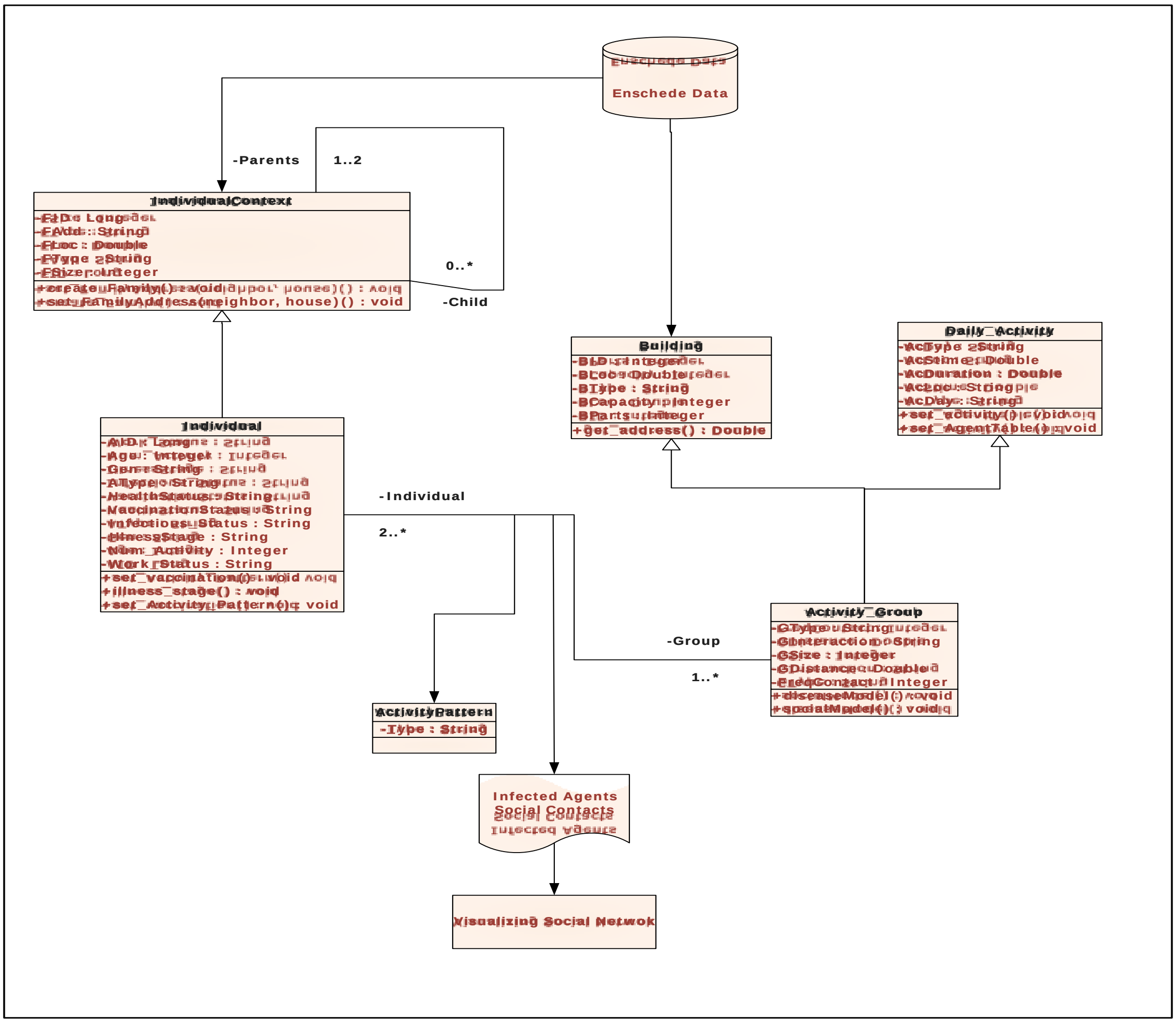


Figure 5.6: UML diagram of the main classes in the model

5.2.1. Synthetic Population (Agents)

In chapter two, the meaning of agents was in ABMs presented and in chapter three, the conceptual agents were presented with the assumptions and structure of creating them. In this chapter, the agents will be created and enabled in the model. To avoid confusion between “agent” and “individual”, both of the terms refer to the same target, which is the real individual in reality.

Creating agents in RepastS can be done in two ways: loading them from a GIS Shapefile or creating them using RepastS directly. The *agent contexts* reads every agent from the Shapefile then load him/her to RepastS, and gives them an appropriate spatial location in the projection.

The agents with the Shapefile may be provided with attributes. To load and use these attributes values within Repast the agent class should contain the setter and getter methods for each attribute in order to read/write values from/to the agent Shapefile (Malleson 2008). However, the set and get method name should start with set/get followed by the same attribute field name with the same spelling and letter cases otherwise RepastS cannot read/write data from the required Shape-file field. For instance, if the Shape-file data has an attribute called *AgentAge*, then the agent class needed two methods called *setAgentAge()* and *getAgentAge* to read and write data in this attribute field of the Shapefile.

Loading agents from the Shapefile might be easier in the term of coding. They did not require methods for creating them or giving them initial geographic coordinates. They would be created only once using GIS tools. Within the creation of them they would be provided with geographical coordinate. In addition, agents could be provided with the required attributes during time of their creation. By using the setter/getter methods agents attribute could be read/written. However, this is very static way and requires a rich data details when they are created. Also positioning them is required within their creation. Loading a Shapefile to RepastS with big number of features add delay to the initializing of the simulation.

The model did not use this type of agents. This is because creating the city population, which consists of 150,000 agents within the Shapefile, will lead to that the initial location of agents/households would be fixed. This would be the correct option to choose if we are sure about the home/family of each individual (having detailed data about the exact address of each family and each individual family members). Nevertheless, we do not have such data. We could manually position each individual, attempting to recreate the cadastre data that we do not have, but as has been stated there are many individuals. These agents need to be created point by point in ArcGIS (with/without attributes) then loaded to RepastS, which will add more delay to the simulation time. Moreover, these agents need to be provided with attributes such as age, gender... etc. i.e., each point should have the demographic data and other required initial values such as vaccination level. Therefore, we choose the second option to create them.

The synthetic population was created by distributing the population totals randomly over the area of Enschede, using the information in census data that was prepared in section 4.1, for instance, totals by neighbourhood, average household size per neighbourhood, average number children per neighbourhood, average number of employed persons per neighbourhood etc. In model.score, *IndividualContext* is the container of the agents (or Individuals). Individual is the agent member of this Context. The container "*Context*" of Individual, which representing the model agents is responsible for creating these individuals/families.

The methods of building synthetic population were discussed in chapter 2. The method that was chosen for this simulation is the synthetic reconstruction approach. This approach is based on Monte Carlo simulation. The procedure of building the synthetic population of the model simply consists of the following steps:

1. Read Enschede census data from the *CSV* file line by line. Each line contains the neighbourhood name, code, number of families of couple parents, number of families of single fathers, number of families of single mothers and 4 age groups of children (0 -4), (5 –9), (10-14) and (15-19).
2. For each line, the household head would be created and assign values are assigned to his/her attributes. In case if a family were couple parents the second adult "partner" would be created too with assigning values to her/his too. These household heads will represent the first families member in this neighbourhood (data record),
3. After creating the neighbourhood households' heads, which represents the families of the neighbourhood, children of that neighbourhood would be distributed over the created families. Based on the assumptions that has been posed in chapter three, the maximum size of each type of families had been taken in account. The children age was assigned with respect to the exact age group size also correspondent to their parents' age, especially the mother's age if the family has the member "mother". The distribution of children over

families is of two stages. In the first stage every family gets one child so that we make sure that we create families with children. The remaining children will again distributed over families as long as the family does not get more than the average assumed.

With the creation of each individual whether individual is a parent or a child, their attributes were assigned. They have three types of attributes: static (demographic attributes), dynamic (changeable attributes based on actions) and reaction attributes. Some of these values would be inherited from parents to children like for the instance, child would live wherever his/her parent live, s/he has the same family ID of parent etc.

The meaning of *static attributes* of an individual is that these values will not change their values during the simulation. For instance, the individual's age will remain static during the simulation time. Of course, if the model was created to represent the development of policies a long a time equals to couple of years, this attribute should not be static. An individual age should be increased by one within each year passed during the simulation, i.e. the static or changed value of a specific attribute is depending on the purpose of the model with respect to the time of the simulation.

On the other hand, some other attributes are dynamic. *Dynamic attributes* change their values during the simulation. The change of these attributes values can have an effect on the simulation. For instance, the disease infects the *sickness* attribute that is used to describe the individual's health status whether or not. This attribute will change when the disease infects the individual. After recovering the attribute will change again to healthy.

The *reaction attributes* includes these attributes, which individual respond to the change that occurs to some of his/her other attributes. For instance, whenever an individual health status changed to become sick, s/he should change the current activity timetable and stay at home until recovered then back to follow her/his the usual activities. The same thing when a child is infected by the disease and becomes ill, one of his/her parent should change his/her current activity pattern and stay at home for nursing. The individual's main attributes can be found in the table 5.1 with the description of them.

Attribute	*Type*	*Default value*
MyID	Static, long	Previous individual.myID + 1
Age	Static, int	Assumption or age group
Activitypattern	Static, array list of strings	--
Sickness	Dynamic, Boolean	False
Friend	Static, list of individuals	Initiate within the initialization of the simulation
Vaccination_level	Double	(1, 0.5 or 0)
Work_status	Static, String	Student, worker Full Time, worker Part Time and Non Worker

Table 5.1: Individual basic attributes

In addition to the above attributes, some others are assigned to individuals which relate to their family structure rather than personal level, for example, *HomeAddress* which identify the home address of the individual so that in the end of the day they go back to the same home s/he came from in the beginning of the day.

The *age* attribute of the individual will distinguish city's individuals into five categories: (0 -3) years old are of type *Infant*, *Child* (4 – 12) years old, *Teenager* (13 – 18) while the adults (age> 18) are of two types: *Worker* (both full and part time workers) and *NonWorker*.

Vaccination level or immunity level another crucial individual attribute as base of it the individual will get the disease or not. This attribute manage the range of spread the disease. In the absence of this attribute, the whole city will be infected or if it set to one "vaccinated" to all model population then no disease will spread. It is a complex individual attribute however; the model used only three values to set for its population base on its data and details of modelling this factor. For the adults this is attributes is closer to be called immunity level.

The model's building block is the individual. These individuals have activity pattern as would be discussed later in this chapter. During the application of their activities, individuals join groups and interact within the level of their groups. An individual in the model has at least one group, which is the group family.

The model does not use the sub-context of family that might be the responsible of combining individuals into family structures. In RepastS, it represents agents as a separate category, therefore; if we want to represent agents of a family, all of its members will be represented as a single point. There will not be individual activity pattern but family pattern, which is unrealistic. In reality, individuals from the same family have different activity patterns. Even though the model still requires to group individuals in unites of families.

In order to combine individuals in family structures, they were provided by a family id. They recognized their family members with the unique family id they have. This *id* helped individuals to know their family members during the simulation. Individuals required knowing who their family members are because both their activity patterns and disease spreading behaviour depend on it.

In addition, for the individuals older than 3 years old they should have a group of friends. The choosing of these group members has some roles assumed in chapter three. For the individuals of age (4 – 18) years old, they choose them based on their age and school. Every one of them chooses two friends from the same age and school. These friends would be added to their memory to keep them fixed and not change their friends every simulation day. This because these friends would become a part of their activity pattern as will be seen later.

For the individuals older than 18 years, they chose their friends differently than their children. Their friends are more variance. The adult friend set consists of a friend from their workplace "colleague", one of his/her children friend's parents and another adult from the same neighbourhood, which might be assigned as the family relative. However, there is no available data of who is relative of who and in order to give the model close representation the reality, one of these friends will be considered as a relative without taking in account what type of relative relationship they are (uncle, grandfather etc.). Generally said, friend set will be recognized by the individual using their memory. They would find their friends and save them in their memory when the simulation is initialized.

By the creation of the synthetic population, the next them is to address them in the city environment where they are going to live and interact.

5.2.2. Loading Enschede City Environment

The city environment consists of the spatial data layer prepared in section 4.2. These layers contain a number of locations where most of human interactions occur. The model projection space is GIS projection space. The building block of the model space is buildings, which have a function, determine which type of interaction/activity should occur there.

The city layers should be loaded to RepastS to build the environment. Previously we stated that agents within the context could not relate to each other to interact unless the context would be provided within the projection, which allow them to act and react.

BuildingContext is responsible for loading the GIS Shapefiles into RepastS projection. The *BuildingContext* is a sub-Context of the *CityContext*. The *BuildingContext* is the container of building objects, which has the geographic projection that is representing the places where the agents were going to visit and join their activity groups. *BuildingContext* provided the individuals with its objects coordinates based on their demand. There is no interaction between the agents and the environment. The interaction is between model’s agents only. Agents are added to the centroid point of the building.

In addition to the data layers which were added to the context's projection. Some other layers were added to the display of the city environment for visualization purpose only. These layers did not load to the *BuildingContext* projection and individuals did not have any information about these layers. By the creation of the population for the model and loading the city environment, the next step would be the presentation of the relationship between the agents and their environment.

The building is used for distinguishing among the activities. It can be said that a building has a function for use that differ from the individual to another based on their type and age. For instance for the individual who has the activity of studying he should address the station of this activity at a school only not other places. The moment the simulation is initialized for running; individuals should explore the environment around and compute the locations for their visiting. This exploration is either done randomly or based on constrains. There will be some terms refer to the place where the individual is going to apply his/her activity such as station, location, place, which have been used in the next sections.

5.2.3. Agents in the environment

The first relationship an individual has with the environment is housing. During the creation of individuals, the household head is given a house coordinate from the same neighbourhood that the individual related to. Later, during the creation of the next members of the same household they would be given the same house coordinate as of their household head. The model chooses the house coordinates randomly. The only consideration is the house should be located in the same neighbourhood of his residential. Many families may be assign the same house. The use of family id is the distinguished between multi families live in the same house building. The house is used for living, also it is the station where the ill individual spend his/her time when becomes ill.

After assigning homes for individuals, they should start looking for their activity stations. Based on individual age and activity type the locations are selected and saved in individuals memory. For the infant individuals of the worker mother, they selected the nearest day care centre, which is close to their residential building. However, the next activity stations of a day would be the same as the parent stations as it has been assumed that infants receive a copy of their parent’s activity. The nearest day care centre will be chosen in the same way as of choosing the schools for children.

For the children the distance is the judge of choosing the activity stations. Even in reality, people choose nearest locations for their children activity for example, parents used to send their children to the nearest school to their house location. The children should look for the nearest Euclidean distance to their houses. The individual of type child who should attend the school will compute the Euclidean distance between her/his house building and the closest school start from 200 m and so one. In case of there would be more than one near school within the same range of distance the child will randomly choose one of them and save the coordinate of the selected school in her/his memory.

The same thing will be applied for children social activity places such as sport centres, music centres, etc. They should select the closest one to their houses. For visiting a friend, the station coordinate will be the friends home coordinate.

Teenagers or adolescents older than 12 years old selected their high schools as well as their social activity stations randomly. They can join any building assigned as high school in the city. In reality, this is applicable. High schools differ from elementary schools in the terms of the most suitability level of secondary schools for children (Oppen-Stuyt 2009). So the model offer random high schools for children of age 13 to 18 to join. The social activity stations also chose randomly but fixed for the next repeated activity.

On the other hand, for the workers who work inside the city they assign a workplace at any position in the city. While workers who work outside the city, they attend a coordinate, which is not within Enschede city coordinate set. The model assumed 2% model population are working outside the city. The available data does not provide information about the percentage of these outside city workers. This percentage is only assumed by the model. However, after exploring the place where individual should visit it fixed in individual's memory to visited again when s/he should apply the same activity another time in the simulation.

The model contains some places for visiting that are not static during the simulation time such as shopping centre. The individual visites different shopping centres every time s/he has a shopping activity. This is also applicable for clubs.

During weekends, people have to go to church, which is depending on the denomination both individual and church then the distance. For instance, even if there is a church close to the agent’s house, if it is not of his/her denomination then this individual will not visited but will look for the nearest same denomination one. For the model, there is no data available about the churches' denominations. Therefore, the determiner of selecting which church to be visited is only the Euclidean distance.

Within the activity location, individuals are differing in their interactions. It has been stated in chapter three that there are two types of interaction within an activity location “full interaction” and “partial interaction”. At least one of these two types of interaction occurs in an activity location. The location and the activity decided which type of interaction existed. For instance, the school station has both types of interactions. If a class in the school is considered then the interaction type is full interaction. All the individuals within this class will fully interact with each other. No boundaries of the maximum number of individuals within the group existed. At the same time, the relation with the other classes will be partially. The individuals of different classes do not interact with each other even they exist in the same place at the same time.

5.2.4. Enabling Agents

Agents were created and the city environment was loaded. They were given a home address. Individuals should have a plan to direct their patterns in the simulation. This plan is the activity pattern and behaviour of individuals in the model. The daily activity patterns largely were determined by the constraints of time and space. A sequence of activities constructs a path in space and time. These activities are defined by location and duration (Ettema and Timmermans 1997). There are two types of behaviour: the healthy individual's behaviour and the ill individual's behaviours.

They are producing different activity patterns. Their behaviour will depend on the values of their attributes *health status* and *vaccination level* (see figure 5.6). There are two types of activities: individual activities and group activities. Within each type of activities, some of them are fixed for every day in the simulation and others are non-fixed.

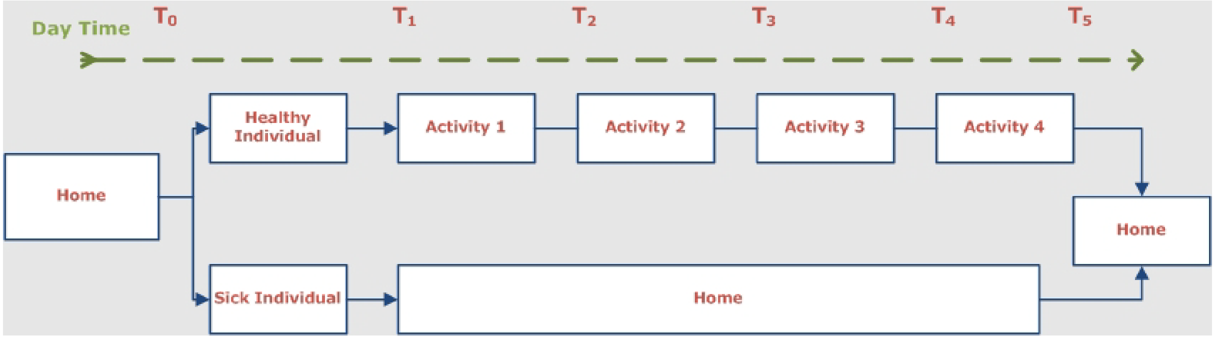


Figure 5.7 : the activity schedule for a simulation day

The activities, which were used in the simulation, are selected from the daily behaviour of the people in reality. These activities were derived from " individuals' basic desires, individual's basic desires, which drive the propensity to engage in particular activities. These basic desires however are not autonomous; they can be "energized" and they can be "constrained"."(Ettema and Timmermans 1997).

Each individual in the simulation has an activity table for the five weekdays (individual activity). The weekend timetable in this model is common among all the family members (group activity). The timetable has been created in a file of the format (.csv). This file is connected to RepastS. The class *ActivityPattern* will read the .csv file and provide the individuals with their specified activities. Each individual based on his/her type extracts the specified activities from this file and create her/his own timetable, which is, stored in her/his memory after the individual created (see figure 5.8).

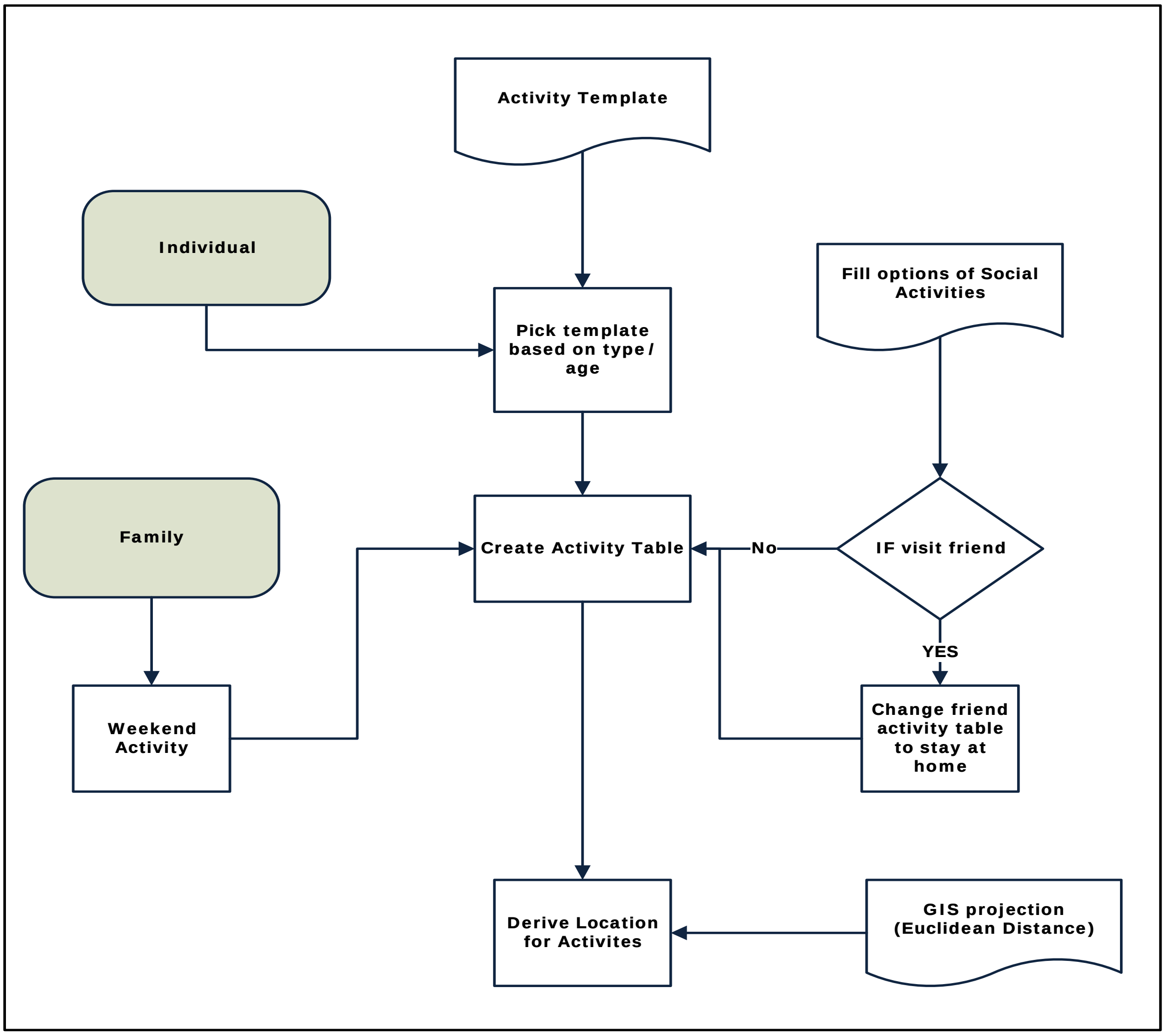


Figure 5.8: Creation of activity table

Every day in the week has fixed activities and non-fixed activities. The non-fixed activities are differing from day to another. These non-fixed activities are the social activities or the activities that are not performed for a special need such as working or studying. They are the activities of entertainment or social purposes such as sports class or visit friends. Nevertheless, for the same day of every week is fixed. In addition, the daytime is fixed for the activities. For instance, everyday at 8:30 the children go to school, at 21:00 should sleep.

The timetable is scheduled in RepastS using the annotations schedule. Agents observing the simulation clock time (tick) whenever they found the time is equal to the next move time they start to leave the current location to the next destination.

When the simulation started to run the individuals will move to their activity locations: students to their schools, workers to their works etc. According to their timetable they make the move based on the time specified for making the move, staying there for a specific duration then after it finished they make the next move and keep moving, staying and making the next move. Every tick in the simulation refers to 30 minutes that means two ticks equal to an hour. For example, at tick 17 means in reality at 8:30 in the morning. The individuals apply their activity and behaviour in method called "*step*". Every individual acting based on his/her own "*step*" method. During the run of the simulation, this *step* method is working for each individual equally. In every tick, every individual check him/herself: "is this tick equal to the time of my next move?".

The individual's movement is disappearing from a place and appearing in another. No regular movement exist using roads network.

The activities that chosen for the model are the most occur in humans' daily life. Some of these activities affect by the economic factor of individual family. For example, the child activity should be based on the parents whether both of them are workers or not. For the worker parents they have the activity of visiting after school hour's care or "*Buitenschoolse Opvang"* which takes care of child of age 4 – 12 years old after school hours and during holidays. It costs the family to pay more per a child. The families of single worker their children assumed not to visit such places rather than they are going to visit their friends.

The idea of choosing such activities to prove the interaction among people led to infecting with disease when they met in the same place at the same time attending the same activity.

5.2.4.1. Agents social interaction

While individuals are applying their moving behaviour, they joined the group of individuals who are sharing the same activity in the same place at the same time. For instance, children in the morning (8:30) moving to schools, join their classmates group, stay there until lunch time then move back home for lunch joining their household members group then again back to school.

One of the models aspect is that individual has another type of relationships in addition to their family relationship, which is friendship relationship. Each individual has a set of friends who are sharing their activities with him or her. After the individuals created and assigned their activity patterns, they have to look for friends based on the assumptions in chapter three. The friends are part of individual's social activity. When the model is setting the coordinates of activity locations, if the activity is visiting a friend the individual look through his/her set of friends. In case if there exist a friend also has the same activity at the same time and day, keep one of them at home and set the other to go and visit his friend. Otherwise, s/he should choose a random friend set his/her activity of the current time and day to stay at home and shift the activity of today to the day when s/he should visit a friend. In the weekend, one of the parent's friends can be assumed as a relative that the family group is going to visit. These social relations between individuals are specified in the *Network Projections* of the *IndividualContext*. The network would consist of nodes (individuals) and their social relations (edges). The projection would store the network or graph relationship between the Context agents. Then it would be easy to inquiry who they are linked and who is linked to them.

The social interaction is a consequence of the activity patterns. It does not occur unless an individual joins a group. This interaction is either partial interaction (a worker is a part of workplace group but interacts with only a part of that group) or fully interaction (a student is a part of class group and interacts with all the group members see figure). This can be clearly seen within the implementation of activity pattern. On the other hand, for the adults, whether workers or non-worker, their groups are sub-groups of the location main group for example a department in a company is a subgroup in the company group. This sub-group distinguished with a specific size. It should not exceed that size.

One of the key factors of disease spread is the social interaction the other factors and the spread of the disease are taking part when the sub model of disease spread is started.

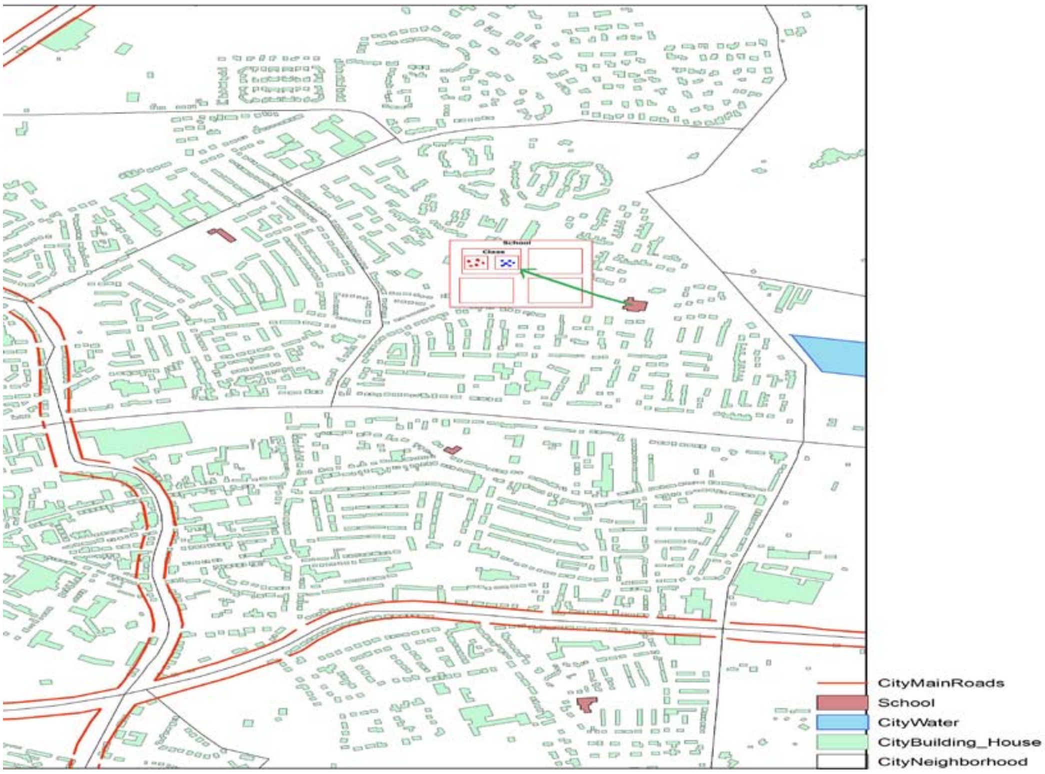


Figure 5.9: Fully and partially interaction within a school building

5.2.4.2. Disease Model

This is the core sub-model of the disease model. This is the part of the model, which provides the perspective about the mechanisms and the way the infectious disease spreads. In this phase of modelling, the disease started in some point and spreads to others. The elements of this sub model are individuals, individual's vaccination level, and individual's social interaction. The disease model and the activity behaviour are related to each other. Spreading the disease prove that the individuals do participate in activities and change groups. At the same time the activity pattern shows how the disease can spread from station to another via an individual which will behave like the vehicle that transfer the disease from station to another.

Individuals are the hosts of transferring the disease from one to another. In the first stage, these individuals are created and provided with required properties (attributes). One of the most important properties an individual has is the vaccination level. This attribute determined whether individual is immune and will not get the disease or is susceptible/non immune and can get the disease at any moment.

The model is populated with the immunity level based on the vaccination data at hand. The implementation determines a vaccination level for each individual based on the assumptions in section 4.1.3. By computing the group size then create the vaccinated ratio of population of that group, we assign a vaccine level to the individual. However, for the adult individuals who assumed to be one of the main sources of mild pertussis, a random ratio of range (20 – 30 %) of total adults in the simulation considered to be susceptible.

Simulation would start with initial infected case(s). The source of their infection is unknown. It had been supposed that the occurrence of the first infection in an unpredictable process due to its dependency on random introductions of the infectious outsiders into the population (Hethcote 2000).

These initial cases will enter the model society based on their activity patterns. For the first three days of the simulation, these initial cases are not infectious yet. For three days are known as the incubation period. Although this period in reality is take around (7 – 10) days but for simplification reasons we assumed only three days. During these three days, no susceptible individual infected as secondary infection.
After four days, the individual becomes infectious. S/he is still running her/his activity table, moving from place to another, joining and changing groups. Two hours after the individual joins a group, the probability of infecting susceptible individuals with the group is increased.
After 4 days of being infectious (day number 7 of been infected), the individual would start to feel ill. That moment s/he should cut the current activity and back home staying there until recovery time.

5.2.5. Next to be implemented

The conceptual design aimed to create a disease model of three stages. The last stage was creating the social networks of infected individuals and visualizing these networks. However, this stage may come in the near future due to the limitation of the current work time.

In addition to not implementing the stage three, there are some other aspects that were not presented in the model. The model population, which is created during running the simulation, remain constant. No new births are added to the simulation as well as no persons are dying during the simulation. The model did not contain any mortality. The model’s environment is constant too. People do not move house or change their activity places such as changing the school or the workplace. Another aspect of using the environment that does not exist with the current model is the movement of individuals.
The model did not present the trips between house location and the activity location. The road even though they were added to the display of the model, but they are not functional to the simulation. Within the actual simulation, the use of road is not defined for the agents.

Although activities vary during the week, the week pattern of an individual is static over different next weeks. The first week is copied to next weeks. In addition, there is no vacation simulated in the model.

The disease model is very simple. There is no reaction when the disease occur except from staying at home. Further reactions are not simulated such as isolating the infected individual when s/he becomes contagious or after became ill s/he will not go to doctor. In addition, the illness stages were not implemented (after become ill until the recovery).
During the epidemic, no preventive measures were implemented. Finally, no controller agent exists such as the “government”, which “wakes up” when the outbreak start preventing people to engage with group activities where the risk is predicted in some places rather than others.

6. Model Results and Verification

The outcomes of the model need to be validating. In this chapter, we will discuss the verification and validation of disease model to judge the results of designing and implementing the model. Verification is the process of checking that the computerized representation of the model matching its conceptual design. It refers to the performance of the implementation code. While validation is referring to the process of checking whether the conceptual design representing the real world (North, Howe *et al.* 2007; Sargent 2007). Verification and validation complete each other.

Simply said, validation checks whether the conceptual model represents the real world then verification checks whether the implementation represents the conceptual model. These two processes are crucial for agent-based modelling. It is important to know the conceptual model and the results of implementing are correct and logically simulate a realistic phenomenon. The three stages of the implementation will be verified to check the model and test it how the model results are close to the empirical data available.

The objectives of the following sections are to present the results of running the simulation and check the models fitness the real data and produce realistic results by verifying and validating the model. The model limitations should be highlighted too. First, section 6.1 present some details about the simulation before getting on presenting the results and analyzing them. In section 6.2, the creation of a synthetic population (model's agents) is presented and discussed, while section 6.3 explaining about the activity patterns. Finally, section 6.4 verified the spread of pertussis disease.

6.1. Starting point

The simulation was run for several times, thirty-six neighbourhoods of the total 70 neighbourhood in Enschede city (around 50% of assumed population to be simulated and 25% of all Enschede population) were input to the simulation. They were selected based on their sequential in the census data but no other reason. Number of agents (individuals) participated in each time of simulation was 34972 agents. They were constructing the model population of the city. Each time of running the model, it took around one hour of the real time. A PC used to run the simulation with RAM equal 4GB, and processor speed equals to 2.4 GHz. The simulation total runtime was equal to 1488 ticks which representing 31 days in real time. It kept constant not changed for every time the simulation was running. The tick in the simulation represents half an hour of real time. The initial infected case was variance between one infected individual and two infected individuals. The simulation at the end of runtime produced some output data files that would be used to verify the models aspects.

6.2. Agent Creation

The model’s agents need to be verified and validated compare with both the conceptual design and input data (census data). The model population is the main actor, the model produced results depend on the population, and its individuals attribute values. Therefore, it required to check the created population.

The model population created using the census data that presented in chapter 4 using the synthetic reconstruction approach. The programming code of the implementation built the city population based on the mentioned approach using connected census data file to read information then create population for the model. The population built consists of individuals structured in families distributed over neighbourhoods. The model could create the exact number of the three types of families.

Source	Tot Families	Tot families of couple parents	Tot families of single father	Tot families of single mother	Children of age (0 – 4)	Children of age (5 – 9)	Children of age (10 – 14)	Children of age (15 -18)	Total number of children
Census data	11290	6794	446	4050	4544	4474	4108	3760	16886
1st Sim	11290	6794	446	4050	6136	4474	4035	2245	16888
2nd Sim	11290	6794	446	4050	6232	4474	4038	2144	16888
3rd Sim	11290	6794	446	4050	6621	4444	2128	3695	16888
4th Sim	11290	6794	446	4050	6553	4435	2207	3693	16888
5th Sim	11290	6794	446	4050	6116	4737	2309	3726	16888

Table 6.1 Statistics of Census data and five times running simulation

Comparing the census data with the results of creating the synthetic population of the model (see table 6.1) some difference can be seen. The difference is in the age groups of children in the model. It can be noticed that for five times of running the simulation, the number of children in the age group (0 – 4) was over estimated. For two times of simulation (Sim 1 and Sim 2), the number of the children in the age group (5 – 9) were constant and for the next simulation times was close to the census data. The number of children in age group (10 – 14) was underestimating in Sim 3, 4, and 5. The same for the age group (15 – 18) but during Sim 1 and 2 it was underestimating.

The reason of that is that the children should have an age value suitable with their parents' age especially the mother's age. Adults' ages are given randomly when the system is initializing the simulation before creating their children. Therefore, when it started to create the children and gave them an age it checked her/his parent's age (his/her mother if she existed) and randomly chose an age for her/him in a range that can be logically acceptable. For instance, a mother with age of 25 cannot have a child with age of 15 years old.

The system is forced to create all the children in such a way that the age groups exist truly as in the reality. Nevertheless, within the random ages of parents the system might not able to find suitable parents for this age of child so it chose another child from another group and so on until in the end the total number of children should exist even if it is not exactly equal to their age group sizes (see figure 6.1).

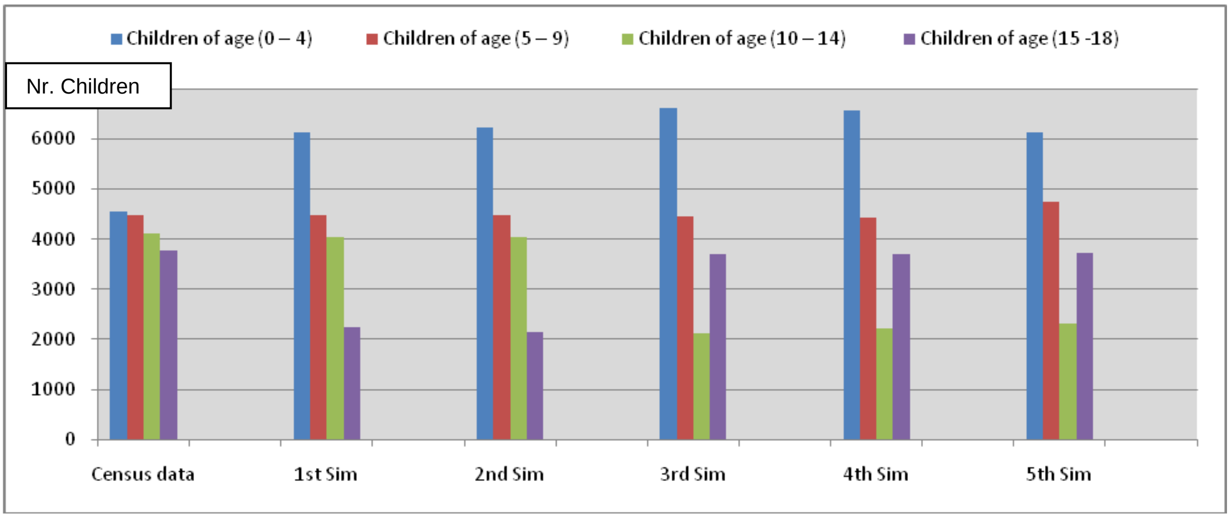


Figure 6.1 The difference between census data and five times of running the simulation

The system also forced the families to have children even if the number of children is greater than the range of a family should have. This type of role leads to have different structures for the model families.

The age attribute is crucial for the model individuals. It specifies the value of the other attributes such as the work status, vaccination/immunity level, individual relationship, type of activity, patterns of activity, severity of illness, and responsibility of household.

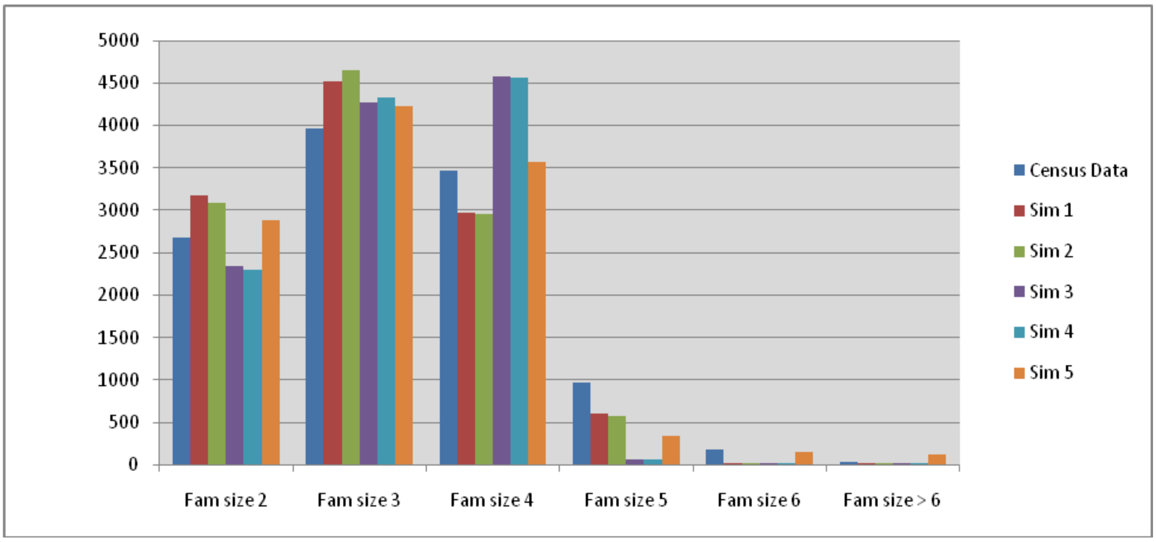


Figure 6.2 The difference between the families sizes in the Simulation and Census data

In reality, families in differ in size. The census data addressed sizes of families starting from two to greater than nine, in our case, the model also provide this variety in families sizes which is different from one neighbourhood to another. The system assigned children to families. It distributed all the children over all the families within same neighbourhood. The system may exceed the average that a family supposed to have based on the assumption. Figure 6.2 shows the difference between the numbers of families of a specific size produced from running the simulation five times compared with the census data. It can be seen that families of size 5, 6 and more than 6 members seem to be an underestimation.

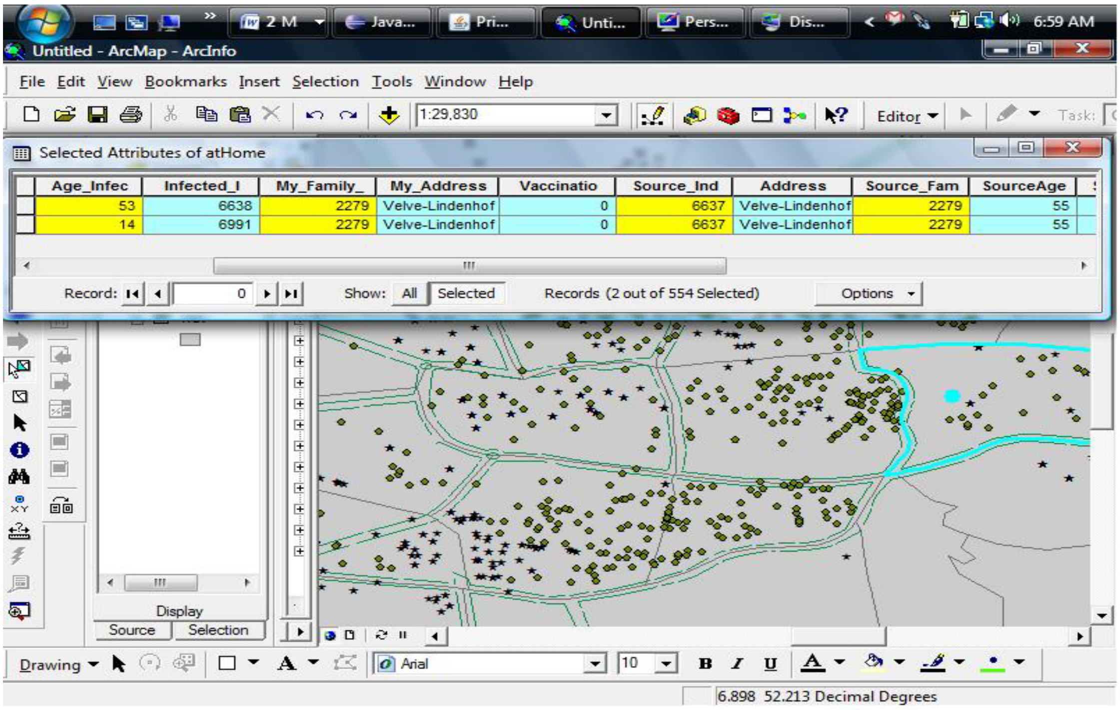
6.3. Loading agents to environment (geography distribution)

During the distribution of individuals over the space in the model, more than one family might be assigned to the same house. In reality, this is acceptable, as buildings with multi-households do exist. Within the model, it did not provide the individuals with any information about the area specified for their families to live in; many families might be added to the same building (centroid point of the house building).

The model combined individuals in families using the family id. . After each household has been created, the family id is increased by 1. The next family would take the incremented family id and so on. To decide whether the correct number of families is created, the maximum number was compared to the total expected number of families.

In this case, the family ID will play the role of the boundaries of the house specified for a family. For example, when two families live in the same building, they are considered two separate groups with no interaction. Although individuals within the same family ID have full interaction (and can transmit the disease), the infection will not automatically spread between the two families (two different family IDs) even when they are sharing the same building. The individual who lives in the same centroid point but with different family ID is not a member for the family. The same story exist in reality if there is a family A with an infected child living in the same building with another family B but with no children or no children of the same age that might have an interaction somewhere, then the disease may not spread to the neighbour family.

This had been proved with the data captured from running the simulation. All transmissions of disease, which is occurred at homes, are from/to the same family members have the same family ID. In addition, by entering the infectious period, which is the fourth day of being infected, individual is at home. The new day in the simulation is the hour 00:00, which is the midnight. That means by that time the individual is at home (sleeping), the transmission started that time to infect the susceptible family members. Only outside homes (during the non-home activities), infected individuals are related to different families (see figure 6.3).



Age_Infec	Infected_I	My_Family_	My_Address	Vaccinatio	Source_Ind	Address	Source_Fam	SourceAge
53	6638	2279	Velve-Lindenhof	0	6637	Velve-Lindenhof	2279	55
14	6991	2279	Velve-Lindenhof	0	6637	Velve-Lindenhof	2279	55

Record: 0 Show: All Selected Records (2 out of 554 Selected) Options

Figure 6.3 Screen shoot of disease transfer between household members

The system randomly chooses houses for families. It does not consider the size of the building or the capacity of how many people can live in this building, but will only consider the "building type". A neighbourhood may contain a lot of buildings of type "house" but the system may assign multiple families to one house and leave the rest empty. This did not cause a big problem in distinguishing families as the system uses an unique family id for each of them. However, it would be more realistic if families are distributed over houses taking into account the building capacity. This can be achieved by assigning an approximate number of families per residential building. Although the data of buildings was provided with the area information, it was not useful because of vertical buildings, which have a limited area but multiple floors. Figure 6.4 shows the distribution of families in a city neighbourhood. From the figure, we can see that families (the dots in the figure) were distributed over only building of type house (in red colour). In addition, the figure shows some of the house buildings remain empty (the red polygons without dots).

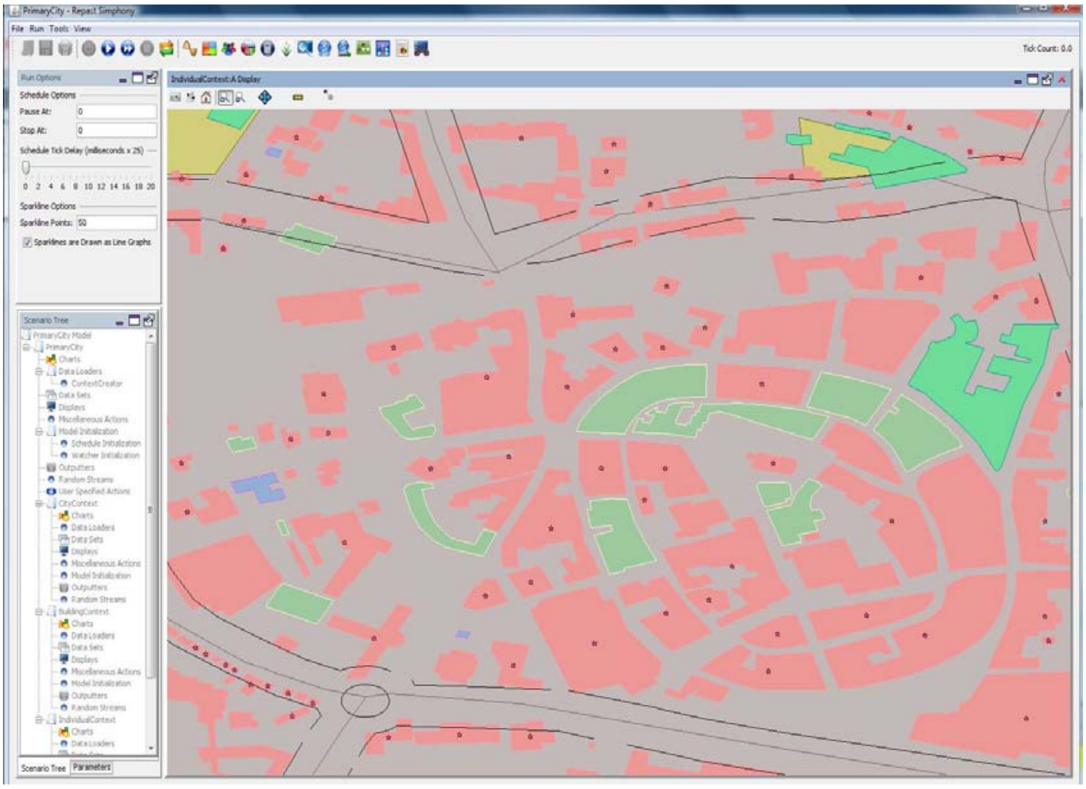
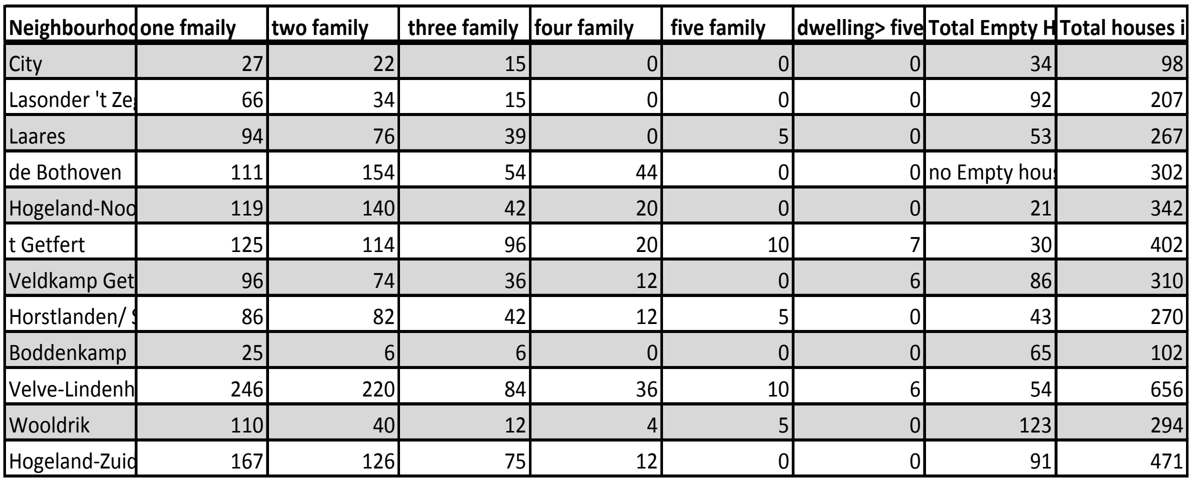

Pause At: 0
Stop At: 0
Sparkline Points: 50
☑ Sparklines are Drawn as Line Graphs
Tick Count: 0.0

Figure 6.4 : Distribution of families over houses

It is good to say that this randomly distribution depends on the number of houses in a neighbourhood. If the number of families per neighbourhood is less than number of houses, it might remain some house building empties. While in crowded neighbourhoods, where the number of houses is less than families there is no empty houses in addition a house building might contain a quite big number of families. The following table shows some statistics capture from running simulation.



Neighbourhoo	one fmaily	two family	three family	four family	five family	dwelling> five	Total Empty H	Total houses i
City	27	22	15	0	0	0	34	98
Lasonder 't Ze	66	34	15	0	0	0	92	207
Laares	94	76	39	0	5	0	53	267
de Bothoven	111	154	54	44	0	0	no Empty hou	302
Hogeland-Noo	119	140	42	20	0	0	21	342
t Getfert	125	114	96	20	10	7	30	402
Veldkamp Get	96	74	36	12	0	6	86	310
Horstlanden/	86	82	42	12	5	0	43	270
Boddenkamp	25	6	6	0	0	0	65	102
Velve-Lindenh	246	220	84	36	10	6	54	656
Wooldrik	110	40	12	4	5	0	123	294
Hogeland-Zuid	167	126	75	12	0	0	91	471

Table 6.2 number of families per a house building in some neighborhoods

Individuals then posed to look for the place where they should live in and construct the family structure. They entered to their environment where they are going to interact with each other. Two things an individual should do during initializing the model: s/he should look for a place to live and places where s/he is going to apply their activities.

Looking for activity locations varies among the individuals based on their age and the function of the location “building”. Although in the reality people choose their locations based on different factors such as distance, money, utilities provided etc, but the model has another aspect in choosing locations for its individual’s activity.

The same role is applied for the infant individuals of the worker mother. They selected the nearest day care centre which is close to their residential building. However, the next activity stations of a day would be the same as the parent stations as it has been assumed that infants receive a copy of their parent’s activity

Worker adults, whether full time workers or part time workers selected their workplace randomly from the provided types of buildings named workplace. However, they may select the shopping centre as a workplace also. Similar to teenagers, adults also chose their social activity station randomly from their city environment.

6.4. Agent relationships (agent’s groups)

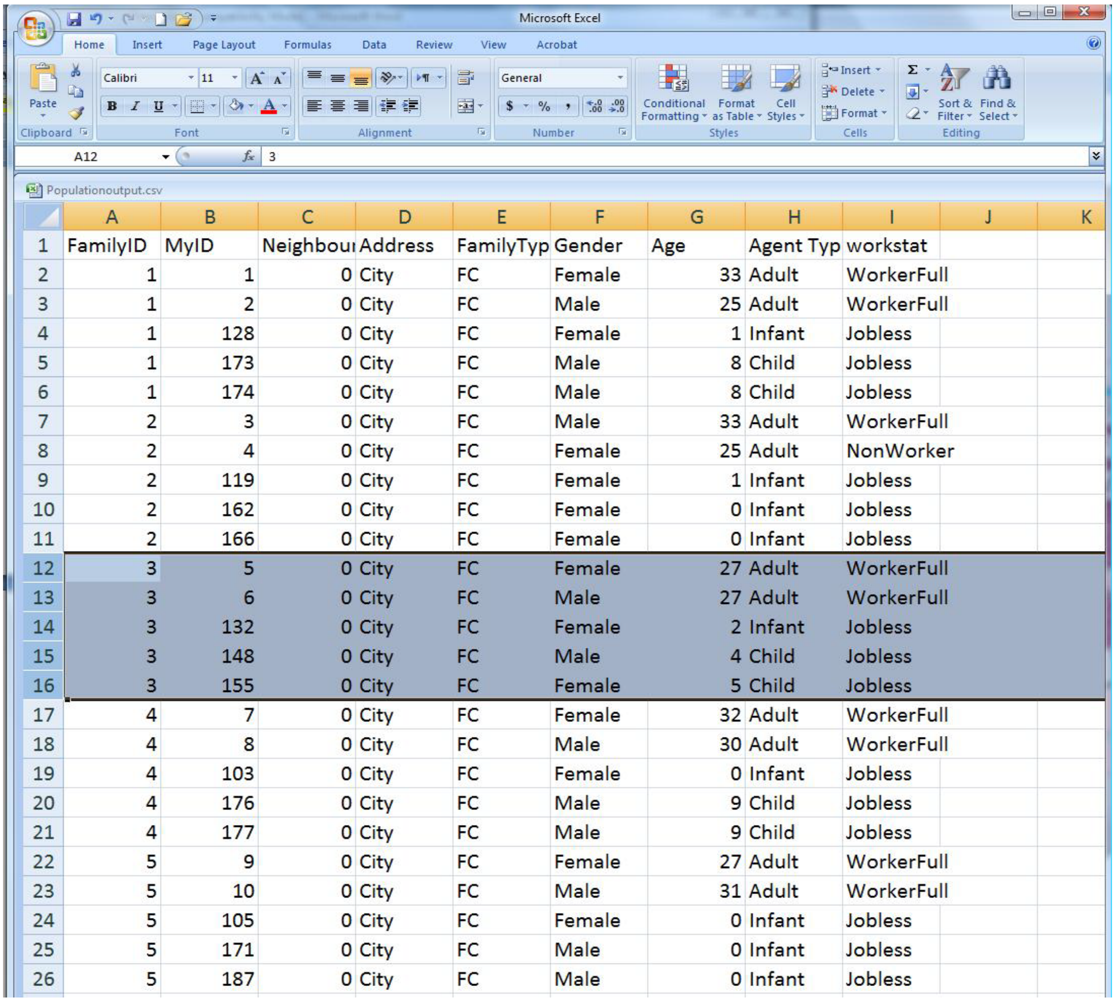
Group behaviour is very important in the simulation and should therefore be checked. The most important groups are the family, friends and the class within a school. The verification of the groups consists of two parts:

- check if the groups are indeed created
- check of the (full) interaction between the group members

To check if families are created according to the specifications in the conceptual design the following tests have been performed:

- Check the captured data from create synthetic population
- Check when the disease spread at location house between the individuals with the same family id.

The earlier check can be done by checking the output data file of the simulation. The group family is connecting its members via Family ID. After sorting the data in the output data file based on the family id, all the members come together who have the same family id. By checking the structure of the family within the type it should be (family of couple parent or single parent) we can conclude whether they exist as a family or not (see table 6.5).



A12 f_x 3

	A	B	C	D	E	F	G	H	I	J	K
1	FamilyID	MyID	Neighbou	Address	FamilyTyp	Gender	Age	Agent Typ	workstat		
2	1	1	0	City	FC	Female	33	Adult	WorkerFull		
3	1	2	0	City	FC	Male	25	Adult	WorkerFull		
4	1	128	0	City	FC	Female	1	Infant	Jobless		
5	1	173	0	City	FC	Male	8	Child	Jobless		
6	1	174	0	City	FC	Male	8	Child	Jobless		
7	2	3	0	City	FC	Male	33	Adult	WorkerFull		
8	2	4	0	City	FC	Female	25	Adult	NonWorker		
9	2	119	0	City	FC	Female	1	Infant	Jobless		
10	2	162	0	City	FC	Female	0	Infant	Jobless		
11	2	166	0	City	FC	Female	0	Infant	Jobless		
12	3	5	0	City	FC	Female	27	Adult	WorkerFull		
13	3	6	0	City	FC	Male	27	Adult	WorkerFull		
14	3	132	0	City	FC	Female	2	Infant	Jobless		
15	3	148	0	City	FC	Male	4	Child	Jobless		
16	3	155	0	City	FC	Female	5	Child	Jobless		
17	4	7	0	City	FC	Female	32	Adult	WorkerFull		
18	4	8	0	City	FC	Male	30	Adult	WorkerFull		
19	4	103	0	City	FC	Female	0	Infant	Jobless		
20	4	176	0	City	FC	Male	9	Child	Jobless		
21	4	177	0	City	FC	Male	9	Child	Jobless		
22	5	9	0	City	FC	Female	27	Adult	WorkerFull		
23	5	10	0	City	FC	Male	31	Adult	WorkerFull		
24	5	105	0	City	FC	Female	0	Infant	Jobless		
25	5	171	0	City	FC	Male	0	Infant	Jobless		
26	5	187	0	City	FC	Male	0	Infant	Jobless		

Figure 6.5 Screen shoot of captured data of synthetic population highlighted a family group

More important is the test when the disease spread at home location and then the infected transmitter and the susceptible receiver are from the same family id. This was discussed in section (6.3).

Besides family relationships, an Individual has friendship relationship. This relationship is of high important in the disease simulation for this model. It is a part of group activity patterns. In order to check this relationship two ways were used:

- Capturing the friend data of individuals.
- Checking the schools if they have classes of student individuals.
- From the activity table of the infected individuals.

The first test was done through checking the captured data of tracking the disease spread. For each infected individual record his/her friend data to the database. Check their family id; check the age for children friends. They should be at the same age. Then checking their vaccination level, if they were immune then they will not occur as a source or destination of the infection.

The second way is via the school. The schools should been checked whether it has classes of children. Each class is of a particular age group. The existence of the class of an age group plus the existence of two individuals who considered friends in the same school is a proof that they are friends.

The third test which is from the activity pattern of individuals. For the time being, the model tracking the activity of infected individuals and their disease spread patterns. Thus, in case of two an infection occur in the activity “visit friend” that would prove that there are two individuals were existed friends and one visit the other then the disease spread. Unfortunately, we could not get the pattern that might back to the fact that if the individuals meet each other at the school and interact then the disease will spread at school. There will not be a chance to spread again at friends home. In addition, if the friend is immune and the infected child visit him/her but there is no disease spread then the data of this visiting will not be captured.

6.5. Activity Patterns (behaviour)

After creating the population, Individuals were provided with their daily activity time table containing information about the activity type and the time of participating in a particular activity in a specific place. By initializing the simulation for running it, individuals compute and save in their memory their activity timetable. In this section, the activity patterns will be verified. It is tested whether individual apply their activity patterns correctly, what type of patterns exist: individual pattern, group pattern or mixed pattern. How did the individual apply their weekdays and weekend patterns? Then the influence of individuals changing activity stations and groups on the disease spread will be highlighted too.

The verification in this stage showed that individuals in the model are following their activity table and joining the correct station and group. Within tracking the pattern of disease spread, the model verified the individual’s activity pattern. The activity patterns proved when:

- The disease spread spatially.
- The disease spreads to people of different age of the person initially infected.
- From one or two initial cases the number of second cases increases.
- Who infected by whom are in the right time, location, participating the correct assigned activity for both.

In order to verifying the participating of individuals in their activities and tracking their patterns of activity as well as the disease spreading while participating the model was planned to build social network showing these patterns and contacts during engaging with groups. However, because of the time limitation of the MSc this could not achieve. Temporally, a solution can provide a proof of these participating and disease diffusion. The solution is to use a GIS tool called “spider tool” which downloaded from the internet (Hawth’s Tool developed by Tony Palmer).

The spider tool helped to connect the point feature representing individuals by polylines features. The spider tool can plot the path of the transition of the disease trough time. The individuals are the ones who are infectious and infect others while attending an activity. The occurrence of a specific individual is only at the time the disease is spreading in the location s/he existed there.

The spider tool connects locations where the disease is transfer from infectious individual to susceptible individual. However, the tool connecting each point in the origin to all the points in the destination whether or not these point were really effect each other. The result would be a very huge file of data containing million connections.

This is a real problem that required to be solved to provide the real connected points only. To handle this problem the produced file of spider was loaded into MS Access to build the query that reduce the unwanted points and keep the really connect ones.

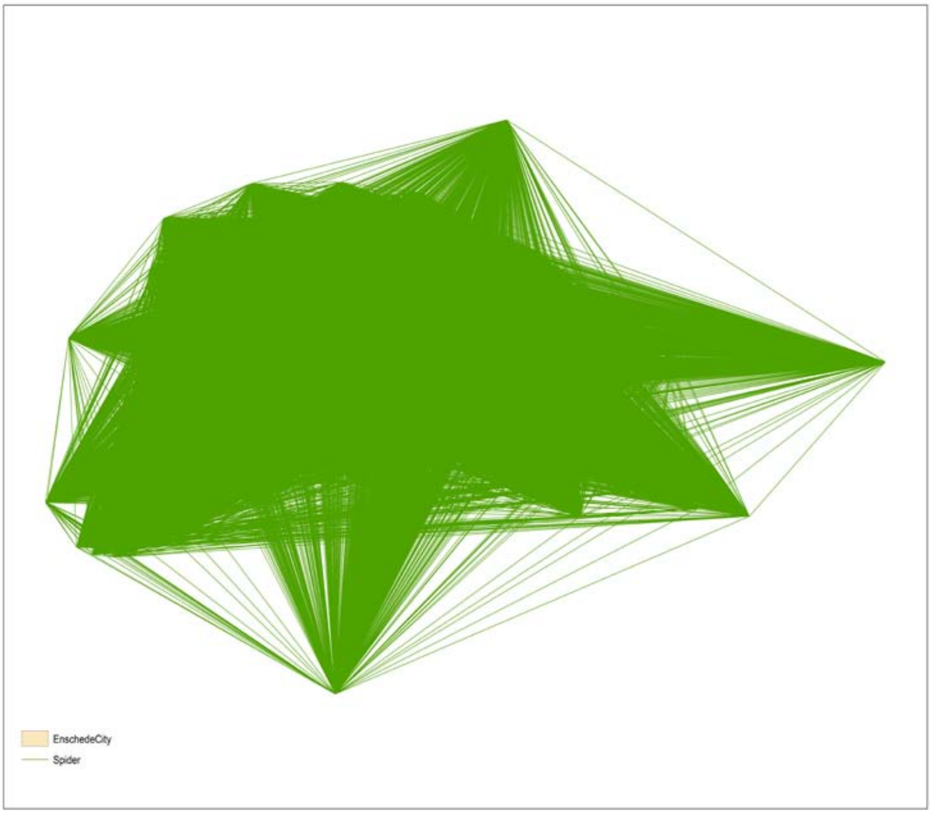


Figure 6.6 : Spider network of (1124) infected individuals

One of the produced file from running the simulation was used. There were 1124 cases of infecting recorded in the Repast file. The file was saved as a spreadsheet then loaded to a geodatabase to be plotted on the map. After creating a GIS layer from the spreadsheet, it was loaded to the spider tool. The spider produced around 1,225,184 polylines (see figure 6.6) which were the output of connecting all the points in source and destination even if they did not have any connection. After querying the spider produced file, the remaining connected points were equal to 991. This is a lower number than the expected 1124 cases.

The reason could be that the tool ignored some because of the big computations it did. It puts them as the end nodes of the connection. They receive a connection without direct a link to another point. The reason of that might be that these points, which represented infected individuals, became infected during the last days of the simulation (26 – 31). Even the ones who were infected on day 26 still have a time to be infectious and spread the disease to others. However, the activity s/he participated in might not include susceptible individuals to transfer the disease to them. Nevertheless, it has been checked if they really did not have a link from (infection source) but some of them has infected some other individuals(see fig 6.7).

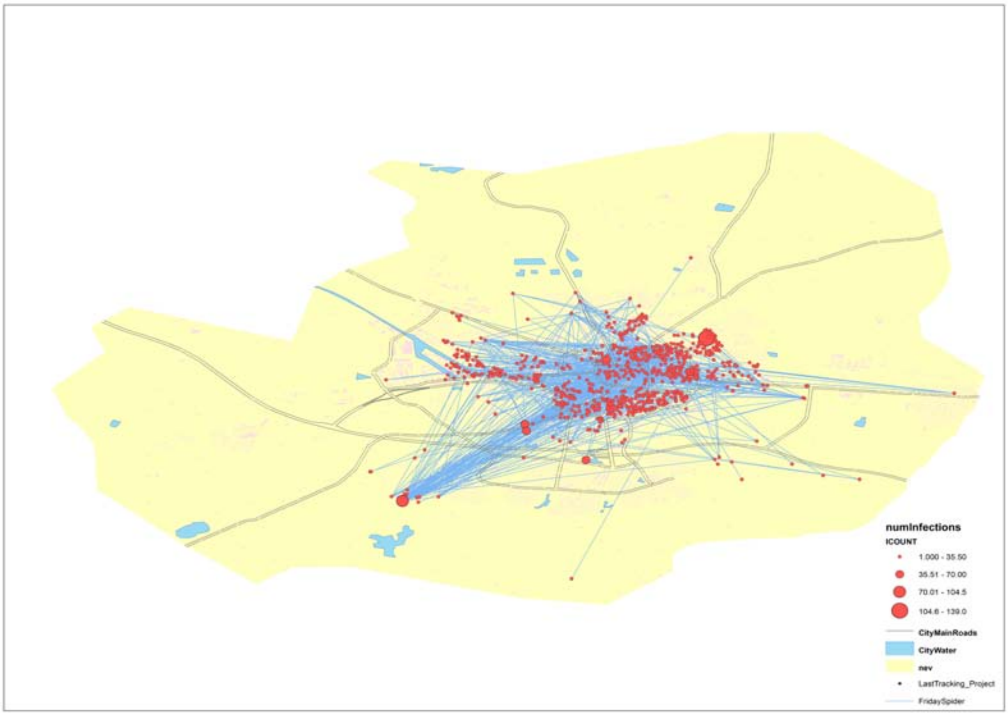
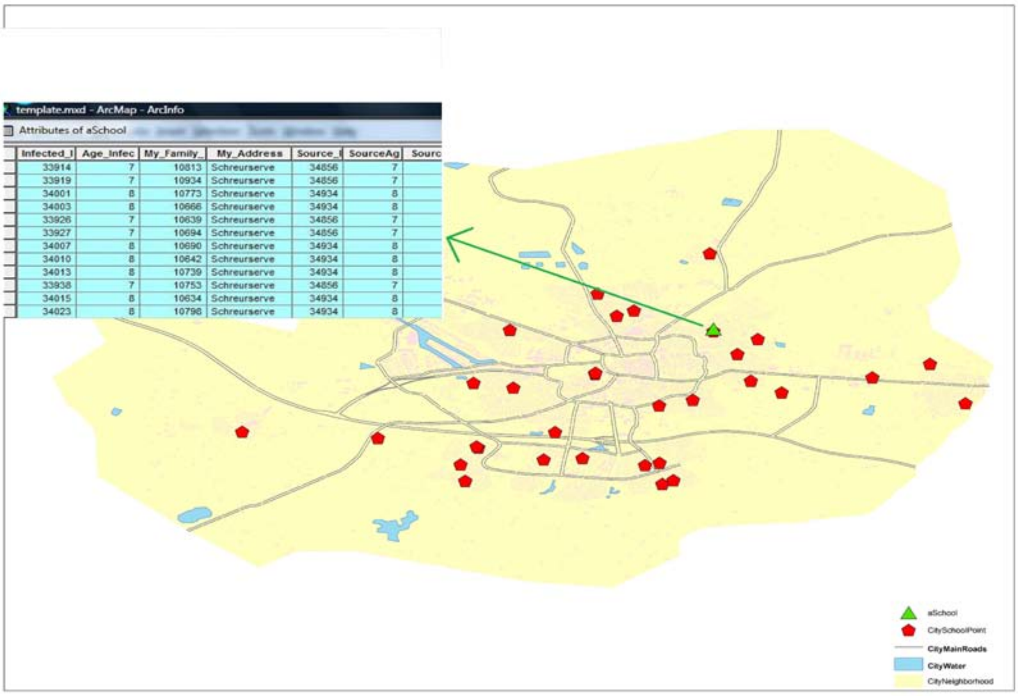


Figure 6.7: The spider network within the infected cases' points

That let me to the conclusion that because they came late in the file somewhere at the end of the file, the tool cannot handle them because of the previous huge computations it already performed. Another explanation is that the amount of computation done by the tool is less than the ones done within the input file. Final explanation might be the misusing of the tool. It could have a trick of producing the correct results, or the data source missed something that is crucial for giving the desired results. The output result of the spider returns 88.16 % of the original points that can provides some confidence about the way of using the method.

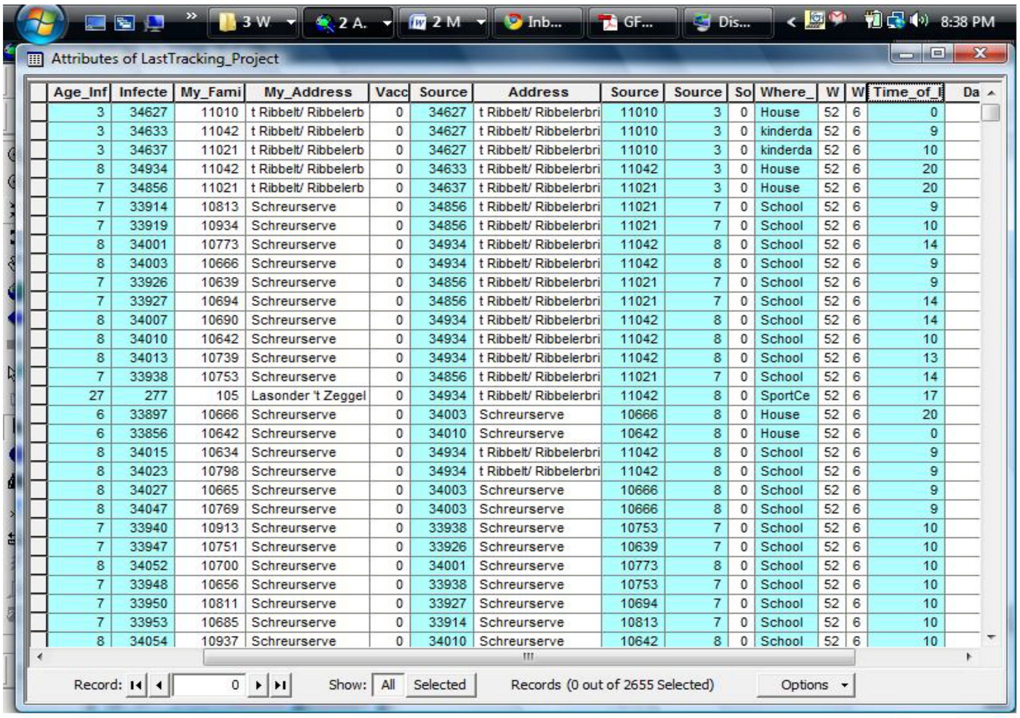
The nice think of analyzing these outputs from RepastS as well as GIS tools, it gives the impression that the model is working properly and the individuals follow what they should do. By selecting, a node from the spider network led to select a couple of point, which had been registered at the same location. The most interesting one is the school. By selecting the school point, it returns some points which are the student individuals attending the school. Then checking “who’s infected who” it clearly shows the class group only infected with their studymate.



Infected_I	Age_Infec	My_Family_	My_Address	Source_I	SourceAg	Sourc
33914	7	10813	Schreurserve	34856	7	
33919	7	10934	Schreurserve	34856	7	
34001	8	10773	Schreurserve	34934	8	
34003	8	10666	Schreurserve	34934	8	
33926	7	10639	Schreurserve	34856	7	
33927	7	10694	Schreurserve	34856	7	
34007	8	10690	Schreurserve	34934	8	
34010	8	10642	Schreurserve	34934	8	
34013	8	10739	Schreurserve	34934	8	
33938	7	10753	Schreurserve	34856	7	
34015	8	10634	Schreurserve	34934	8	
34023	8	10798	Schreurserve	34934	8	

Figure 6.8: Infected students from different classes in a school

As been stated, to verify the activity patterns and behaviours of model agents and prove they are working properly the simulation tracked the patterns of disease spread. Individuals are the vehicle of transferring the disease from person to another and from place to another. Figure 6.6 is of the data captured from running the simulation. The highlighted fields is about who infected whom, where, when, are they from the same location group.



Age_Inf	Infecte	My_Fami	My_Address	Vacc	Source	Address	Source	Source	So	Where_	W	W	Time_of_I	Da
3	34627	11010	t Ribbelt/ Ribbelerb	0	34627	t Ribbelt/ Ribbelerbri	11010	3	0	House	52	6	0	
3	34633	11042	t Ribbelt/ Ribbelerb	0	34627	t Ribbelt/ Ribbelerbri	11010	3	0	kinderda	52	6	9	
3	34637	11021	t Ribbelt/ Ribbelerb	0	34627	t Ribbelt/ Ribbelerbri	11010	3	0	kinderda	52	6	10	
8	34934	11042	t Ribbelt/ Ribbelerb	0	34633	t Ribbelt/ Ribbelerbri	11042	3	0	House	52	6	20	
7	34856	11021	t Ribbelt/ Ribbelerb	0	34637	t Ribbelt/ Ribbelerbri	11021	3	0	House	52	6	20	
7	33914	10813	Schreurserve	0	34856	t Ribbelt/ Ribbelerbri	11021	7	0	School	52	6	9	
7	33919	10934	Schreurserve	0	34856	t Ribbelt/ Ribbelerbri	11021	7	0	School	52	6	10	
8	34001	10773	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	14	
8	34003	10666	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	9	
7	33926	10639	Schreurserve	0	34856	t Ribbelt/ Ribbelerbri	11021	7	0	School	52	6	9	
7	33927	10694	Schreurserve	0	34856	t Ribbelt/ Ribbelerbri	11021	7	0	School	52	6	14	
8	34007	10690	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	14	
8	34010	10642	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	10	
8	34013	10739	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	13	
7	33938	10753	Schreurserve	0	34856	t Ribbelt/ Ribbelerbri	11021	7	0	School	52	6	14	
27	277	105	Lasonder 't Zeggel	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	SportCe	52	6	17	
6	33897	10666	Schreurserve	0	34003	Schreurserve	10666	8	0	House	52	6	20	
6	33856	10642	Schreurserve	0	34010	Schreurserve	10642	8	0	House	52	6	0	
8	34015	10634	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	9	
8	34023	10798	Schreurserve	0	34934	t Ribbelt/ Ribbelerbri	11042	8	0	School	52	6	9	
8	34027	10665	Schreurserve	0	34003	Schreurserve	10666	8	0	School	52	6	9	
8	34047	10769	Schreurserve	0	34003	Schreurserve	10666	8	0	School	52	6	9	
7	33940	10913	Schreurserve	0	33938	Schreurserve	10753	7	0	School	52	6	10	
7	33947	10751	Schreurserve	0	33926	Schreurserve	10639	7	0	School	52	6	10	
8	34052	10700	Schreurserve	0	34001	Schreurserve	10773	8	0	School	52	6	10	
7	33948	10656	Schreurserve	0	33938	Schreurserve	10753	7	0	School	52	6	10	
7	33950	10811	Schreurserve	0	33927	Schreurserve	10694	7	0	School	52	6	10	
7	33953	10685	Schreurserve	0	33914	Schreurserve	10813	7	0	School	52	6	10	
8	34054	10937	Schreurserve	0	34010	Schreurserve	10642	8	0	School	52	6	10	

Record: 0 Show: All Selected Records (0 out of 2655 Selected) Options

Figure 6.9 : database captured from running a simulation

It can be clearly seen that for the individual who infected at home, his/her source of infection was one of his/her family members. Both of them have the same family id and locate in the same location. While for the individual who got the disease from the school, their family address is different as well as the home address, but they are from the same class (has the same age and school). In addition, the time of the activity is shown in the figure telling the time of the activity. By comparing the output data of this table with the input template of the activity tables, individuals were participating in their activities on time.

Figure 6.10, it is the tracking of pertussis transfer in day 22 of one of running the simulation. The colourful lines (produced by the spider tool) are representing the path of disease transmission from an infectious individual to the susceptible individuals. It connected the infected individual (source) location with the susceptible individual (destination) location that would infect another individual (becomes a source for the next). If we follow the purple line in the figure 6.10, which is connecting an individual at home that infected another one would go to school (green line with label 9) infecting one of his classmates then when the later one back home s/he would infect his/her brother/sister. The brother/sister would go to the “kinderdagverblijven” and infecting an infant there and so one. The lines with ends (not connecting to another line) means whether the next day of infecting not coming yet or will finished there and infected individual will not have further simulation time to infect another individual.

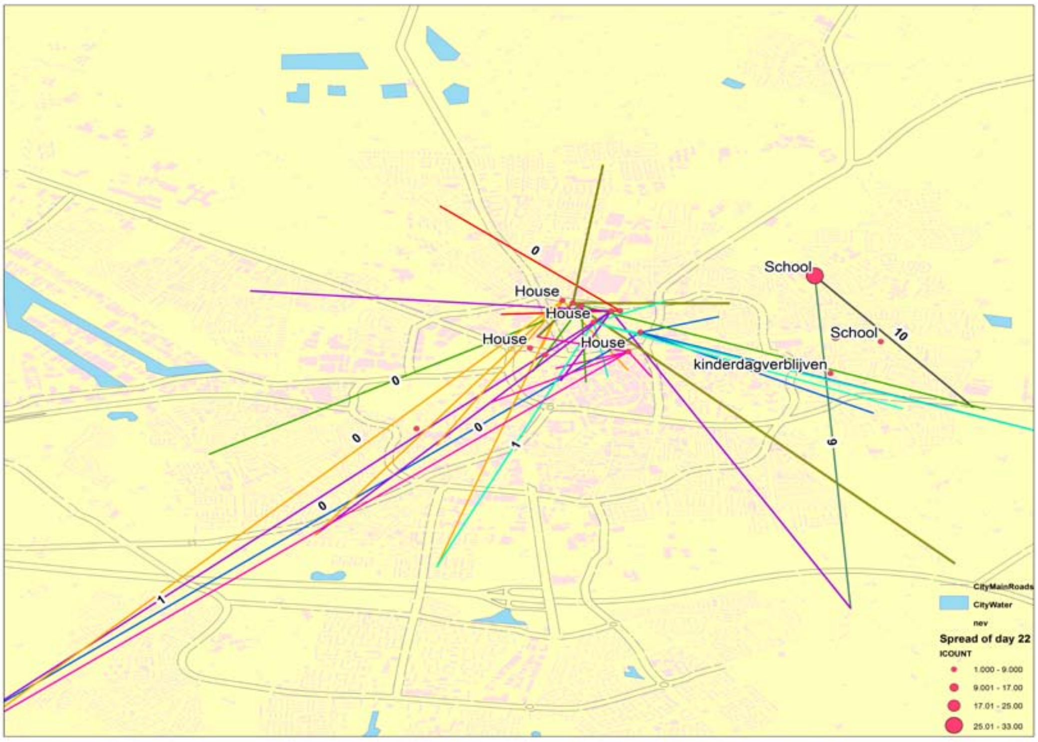


Figure 6.10: Capturing spider network of a day of disease diffusion

For the figure 6.8, this is tracking a weekend activity of family groups. If we make a look at the green lines in the figure, we can clearly see that the infected individual in the “Religious building” are connecting the locations where they come from and where they are going to infect.

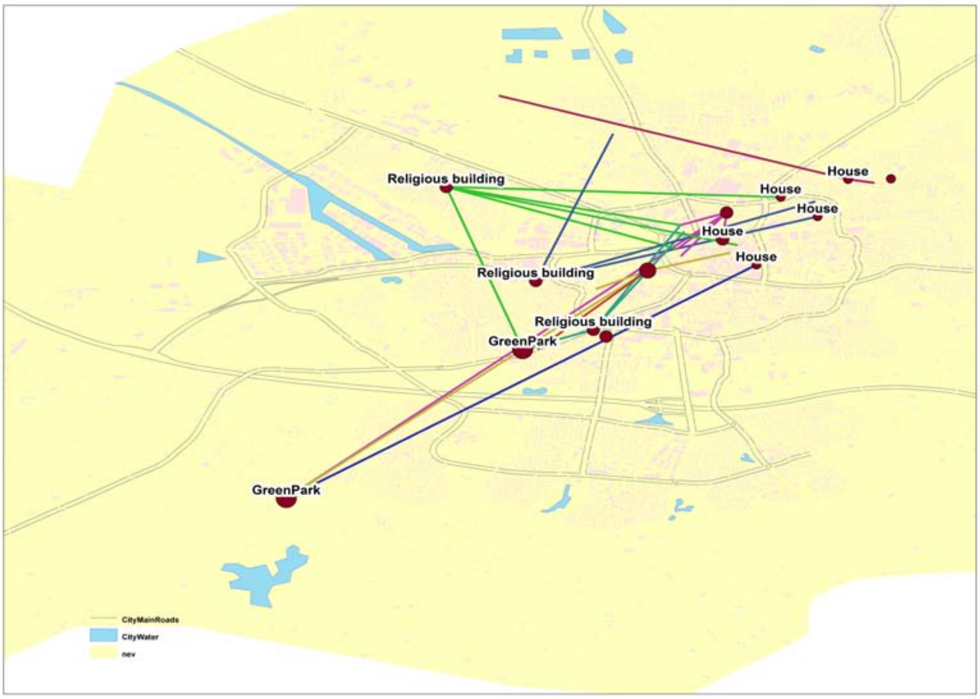


Figure 6.11: a weekend disease spread based on activity

It can be conclude, the activity patterns (agents' behaviours in their space-time) and the spread of disease are nested process, each effect on the other and each can be a proof for the other. Together, these two models provide the meaning of interaction between the model's sub models.

6.6. Disease Model (event and reaction)

Verifying the pertussis disease model is quite complex for many reasons. In additions to the complexity of the model itself, pertussis is one of the complex infectious diseases. Simulating such a disease required rich data about disease control measures such as vaccination coverage, contact rate and information about the disease itself. Most of the researches of pertussis show the complexity of the disease structure (i.e. its bio structure), including the immunity level of an individual against this disease.

The simulation has been run for several times, each time there were one or two different initial cases (infected individuals). These initial cases have been chosen from different group of ages and sometimes there where combinations of starting with two initial cases related to different age groups. The reason of choosing different initial cases was the different activities of these different types of individuals have, the vaccination level was distributed randomly within the specific proportions over the population. These diversities of choosing initial case(es) led to different results and patterns of spreading the disease. This selection will help to explain the influence of these two aspects on disease spreading. The initial combinations of two initial cases for each run are shown in table 6.2.

Simulation time	**1st initial case**	**2nd initial case**
Sim 1	Individual (0 – 3) years old	Individual (0 – 3) years old
Sim 2	Individual (0 – 3) years old	Individual (4 – 9) years old
Sim 3	Individual (0 – 3 years old	Individual (10 – 18) years old
Sim 4	Individual (0 – 3 years old	Individual Non-Worker (18<) years old
Sim 5	Individual (4– 9) years old	Individual (4 – 9) years old
Sim 6	Individual (4– 9) years old	Individual (10 – 18) years old
Sim 7	Individual (4 – 9) years old	Individual (18 <) years old
Sim 8	Individual (10 – 18) years old	Individual (10 – 18) years old
Sim 9	Individual (10 – 18) years old	Individual (18 <)years old
Sim 10	Individual (18 <)years old	Individual (18 <) years old

Table 6.3 : Initial cases infected with pertussis

As a result of ten times of running the simulation with different combinations gave some results. These results are shown in table 6.3 and 6.4. The first table (6.3) shows the percentage of unimmune individuals in the population and the total number of infected individuals per running the simulation. The distribution of the total number of infected individuals over the age groups is shown as well.

Simulation	**Total population**	**Percentage of un-immunes**	**Total number of infected**	**Infected of age (0 -3)**	**Infected of age (4 – 9)**	**Infected of age (10 -18)**	**Infected of older age**
Sim 1	36639	23.4	512	4	0	0	507
Sim 2	36638	25.38	2639	8	140	1459	1031
Sim 3	36638	16.89	1807	12	65	709	1021
Sim 4	36638	18.63	3050	34	0	0	3061
Sim 5	36638	20.07	4949	24	207	2051	2217
Sim 6	36638	17.85	4412	28	179	2072	2115
Sim 7	36638	14.89	3857	32	189	717	2919
Sim 8	36638	25.92	6635	28	85	3817	2705
Sim 9	36638	17.40	1063	86	0	106	870
Sim 10	36639	14.36	3406	12	39	1987	1368

Table 6.4: immunity level and the infection distribution over age groups

It can be clearly seen that each time of simulation return back different results. The analyze of this differentiation in the results might back to couple of reasons; it might be the vaccination/immunity level among the individuals in the population was high as well as the activities they attend. They might engaged with a group with no susceptible individuals or contain high ration of susceptible individuals.

The vaccination/immunity level of the model population were assigned by computing the group size then create the vaccinated ratio of population of that group, we assign a vaccine level to the individual. However, for the adult individuals who assumed to be one of the main sources of mild pertussis, a random ratio of range (20 – 30 %) of total adults in the simulation considered to be un-immune susceptible to get this disease and transmitted to other individuals.

In addition, the children distribution over families has a great effect on the disease spread too. Wendelboe, Njamkepo *et al.* (2007) showed that parents, siblings and non-household close contacts are important source of pertussis diffusion to infants. They provided robust evidence that household members are in charge of 73% - 82% of pertussis diffusion to infants.

Infants in the simulation had a copy of their parents' activity (specially the mother if she existed) unless the parent is a worker that time, the infant went to day care centre. Infants of age (0 – 3) were considered to be without friends and the only contact these individuals have is with their family members. So in the case of not having brothers/sisters older than 4 years their infectious disease only transfer to their parents who spread it to other adults. While in the other hand, the time that a child/teenager was the initial case then the disease spread to other children/teenagers either via the school or via the social places. Here the effect of groups and activities the individual participated and engaged can be see it clearly.

The number of infected individuals could be considered high due to the rapid time of increasing the cases and the absence of individual reactions (stay at home, go to doctor etc.) see figure 6.12. However, if we see the number of cases including the infants or small children who are the most group in risk we can say it is not that much high to be unacceptable (see table 6.4). In addition, pertussis in adolescents and adults is mild so they cannot feel that they are ill. These situations were existed in reality too and as presented in the previous chapters it is one of the reasons that made pertussis remain endemic. Even with comparing the results with the empirical data, the results are remaining acceptable if we consider the severity is strong at the infant levels.

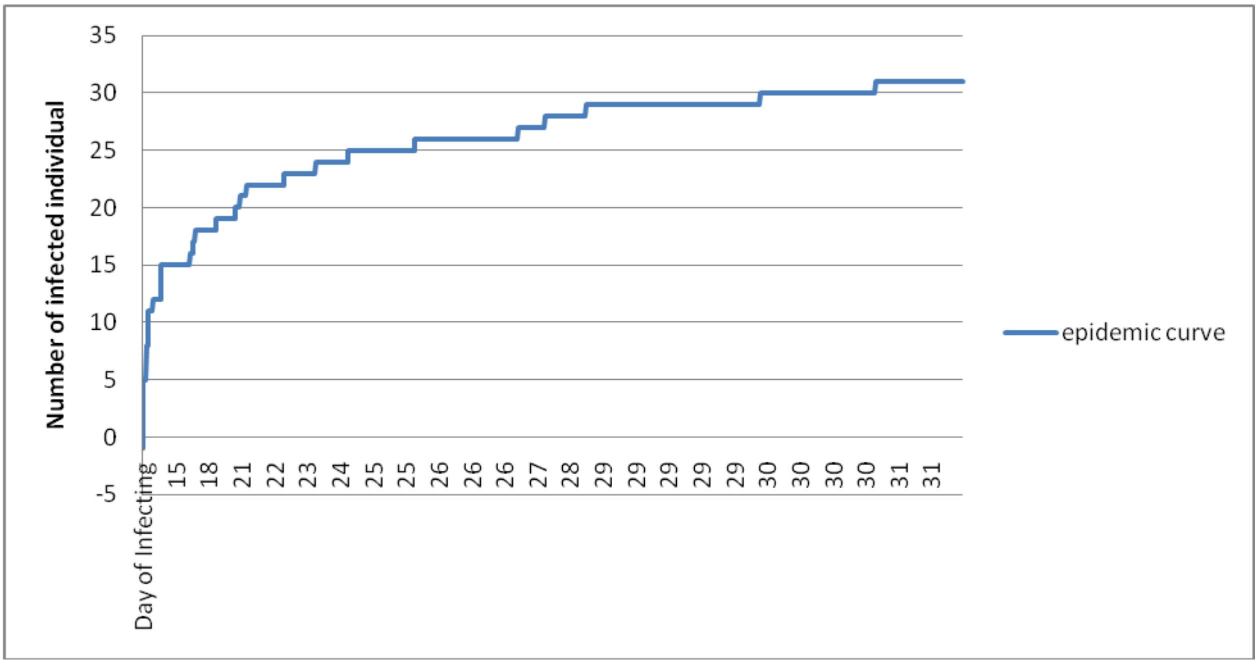


Figure 6.12 : infections occur during the simulation time

Each infectious disease has what is called the reproductive number (R_0). Reproductive number is defined to be the average number of secondary infectious cases produced by infectious cases when introduced to a susceptible population (every one supposed to be susceptible for getting the disease) (Hethcote 2000). An infection can get start if and only if $R_0 > 1$ based on many deterministic epidemiology models. Thus, R_0 is considered as a threshold, which determines when an infection can infect a new host population (Hethcote 2000).

Based on the model presented by (Hethcote 1999a) not all infective individuals are not equally in effecting new infectives. For his models (Hethcote 1997; Hethcote 1999b) and he used the values of $R_0 = 5.4$ for the first one and $R_0 = 3.7$ for the second one. The first model simulating the pertussis with the absence of booster vaccination, while in the second model he introduced booster vaccination to the model (Hethcote 2000). For the current model, the system will choose for the infectious individual a random value of reproductive infection that not exceeds the value of R_0. If the system assumed that the model will be run for non booster population so R_0 will set to 5.4 or if the population supposed to be booster vaccinated then $R_0 = 3.7$.

In the case of the existence of control measures of the disease which is in our case is the existence of vaccination, then R_0 is not used but the control reproductive number R_c is used. R_c is defined as the effective reproductive number when a control measures are existed. Within the presence of control measures R_c is derived from the vaccination level of the population as the following equation:

$$Rc = R_0 (1- h f)$$

Where f is the ratio of population that have been vaccinated or the vaccination coverage of the population. While h is the ratio of completely vaccinated individuals in the population. by computing the Rc of the model population it will be the maximum number of susceptible individuals that an infected individual may transfer the disease to per each station (activity group).

The moment the infectious period of the infected individual started, within the group s/he joined s/he will scan the group looking for susceptible individuals counting them if they are more than the number that selected for her/him randomly, s/he infects only the range available. In case that the susceptible in the group are less than the assigned number then s/he infected only that available number of individuals.

The susceptibility of an individual determined by his/her level of immunity or vaccination. If the vaccination level (for individuals age < 18) or immunity level (individuals age > 18) is equal to zero and they existed in a group containing an infectious individual then they got the disease. After receiving the disease again like the initial ones, they have incubation period of three days sequenced by four days of infectious then illness and staying at home.

The moment the individual got the disease, s/he remains applying her/his activity pattern for a while. Individual cannot feel ill immediately after getting the disease. During this time, individual participate in his/her activities. After four days, s/he become infectious but still applying his/her daily activities; interact with the other members of the groups, which they engaged with.

By using the space-time cube tool to visualize the output data of the simulation, we got the results of spreading the disease from one/two initial case to around thousand of second cases belong to different neighbourhoods and locations (see figure 6.11).

The space-time cube tool used the coordinates and the date of occurrence of the infections from the data captured from running the simulation, the background map of Enschede used to provide the visualization of the diffusion on the city level.

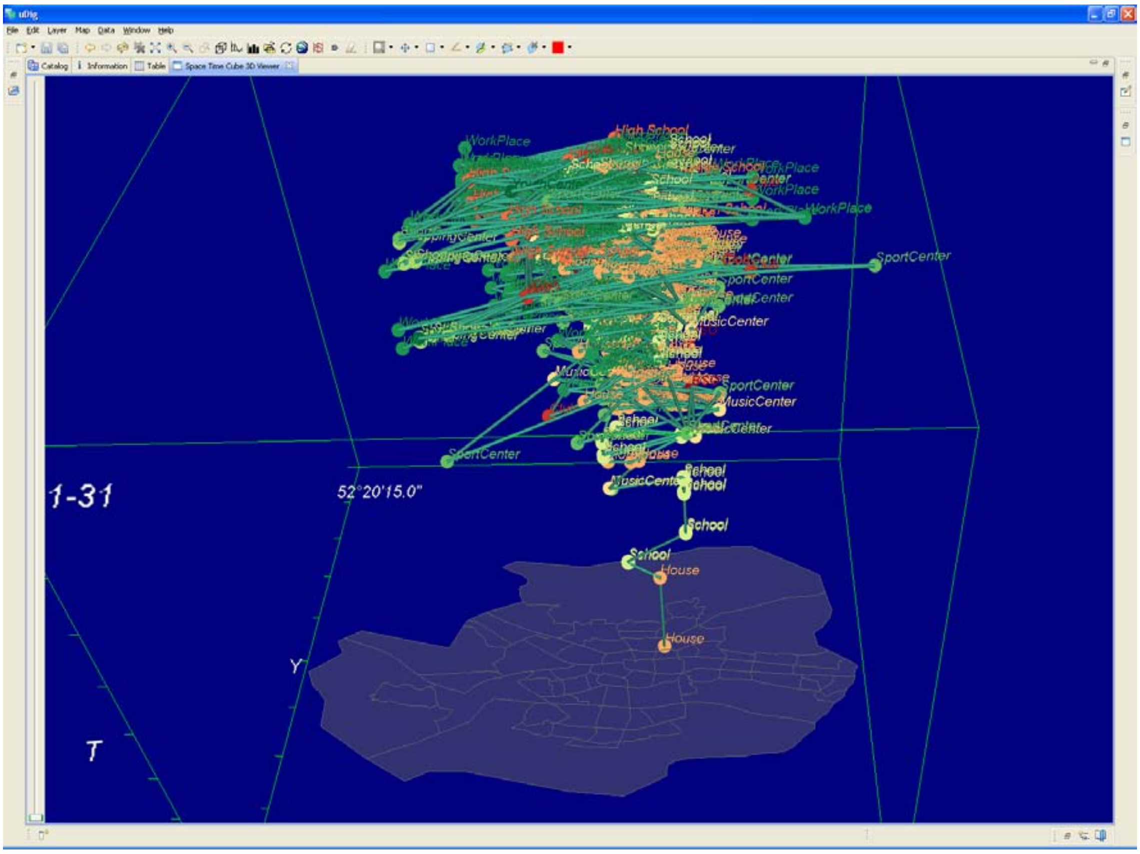


Figure 6.13 : Screenshot of the space-time cube showing the pattern of disease diffution

The next screenshot (figure 6.14) shows the time development of the diffusion. It illustrates how the initial case is non-infectious for a while then started to spread to few number of individuals.

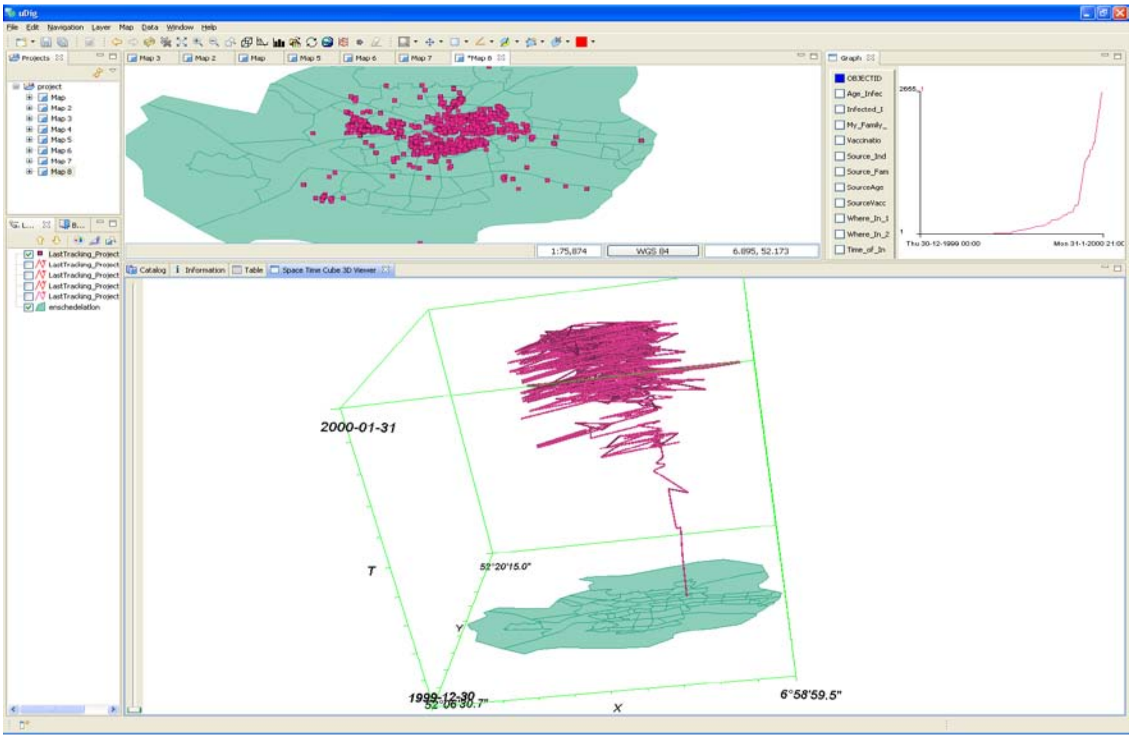


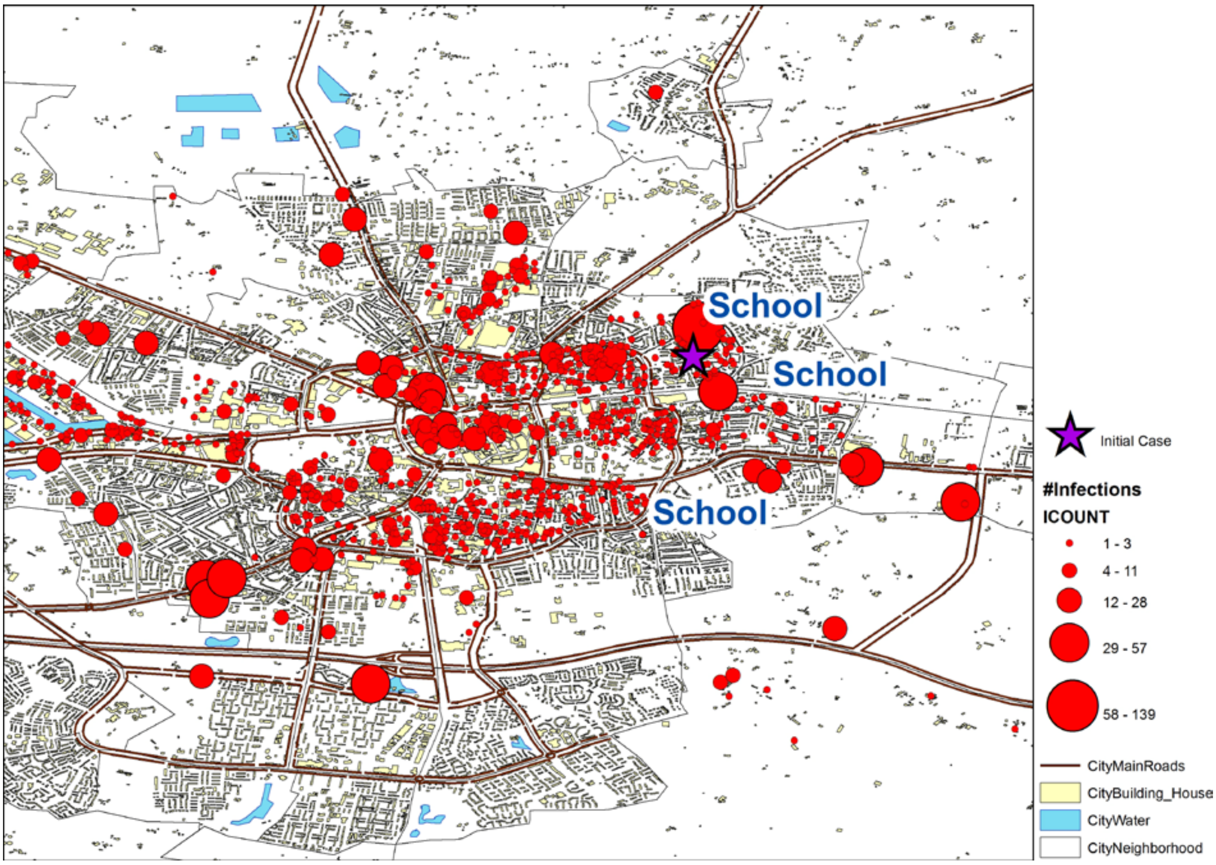
Figure 6.14: screenshot of the evolution of the disease by the time

On the other hand, table 6.4 shows some of the captured data of the percentage of infection occur in activity stations. The main stations are house, school (both elementary and high schools), workplaces and social places, which include every other place.

Stations	House %	School %	Workplace %	Social places %
Sim 1	17.57	0	37.1	44.92
Sim 2	19.85	46.63	11.41	19.62
Sim 3	20.96	32.02	15.21	28.53
Sim 4	24.35	0	44.17	31.45
Sim 5	20.87	43.4	18.63	17.09
Sim 6	22.37	40.52	22.21	14.86
Sim 7	25.20	16.95	35.59	23.07
Sim 8	18.76	49.02	18.23	13.95
Sim 9	21.07	9.31	16.83	52.58
Sim 10	19.14	47.21	12.71	20.93

Table 6.5: percentages of disease occurrence in locations

The table 6.3, gives the proportion of the cases occur in the main locations of the simulation, three of them are fixed activities and been visited every day. These fixed places has quite high proportion of the spreading of the disease.



After 4 days of being infectious (day number 7 of been infected), the individual would start to feel ill that moment s/he should cut the current activity and back home staying there until recovery time.

During staying at home, s/he considered being in the illness period which comes after infectious period. However, in reality the individual remain infectious even in the time of illness, as been explained in chapter 2.

After one week, which was considered, to be the infectious period of the disease the individual should stop applying the usual activities and stay at home. After getting the disease and staying at home, the individual should wait there until recovery moment. After recovered from the disease two main things happened to the individual: the individual should back to the usual activity pattern, and the vaccination level would set to immune level (or 1). In this case, if there is an infected individual (in his/her infectious period) joining the activity where the recovered individual is s/he there the individual will not get the disease again.

The model has a couple of limitations, however, it can prove that it is work and the disease is spread base on the attendance of the infected individual in an activity, joining his/her group there and then spread the disease to the group members who were un-immune against pertussis.

Before analyzing the above results, the following points should be in account:

- The disease spread in the model depends on the duration of contact, the susceptible immunity level, and the reproduction control of the disease.
- Being in contact means both infected and susceptible individuals are at the same location and engaged with the same social group.
- The immunity level of the whole population is high (around at least 75%)
- For the infected individual, there will not be any recovery.
- No population reaction against the epidemic occurs in their city.
- The other controllers are absence except the vaccination.
- The simulation is for one month (31 days)

Of course there are still remaining issues that required to take into account such as the limitations. Some of them related to the data and tools used for the simulation and others related to the research itself. More detailed data required for the model. This data includes details about the creation of the population. The simulation currently has lake in distributing children over families with appropriate age that avoid the duplication of more than one or two children have the same age. Data about the immunity level of the city is required. This data can help to build more accurate synthetic population for the model.

The simulation in the level of groups should be found and connected to the simulation. The current model assumed by sharing time, space and some other attributes such as the age component a group for the individual that has the same purpose of participating a specific activity.

The reaction to the occurrence of the disease is required. For instance, when a child becomes ill, one of his parents should cut his/het activity pattern and stay at home nursing the child until the time of recovery.

A controller required for the simulation. The controller is an agent who wake up when the cases of the disease reach a specific threshold threat an epidemic occur and some specific procedures should run. The procedures may be giving antibiotic for the rest of the infected individual household members, close schools to avoid the wide spread of the disease.

Within the infected with disease the severity of the illness should be simulated. As it is effect on the patterns of activity. However, even with meld infection, the individual remains infectious and spread the disease to others.

Within the simulation itself, the illness model need to be create and connected to the model. For time being, if the individual becomes ill, s/he should stay at home until the last moment in the simulation. No recovery from the sickness provided.

Finally, it will be very fine to apply the simulation with another dataset or another disease to check the flexibility provided within the current model.

7. Conclusion and Recommendations

7.1. Conclusion

The objective of this research was to create an agent-based disease model simulating the spread of pertussis in the Enschede region. Important aspects of the simulation are the recreation of the population, the activities the population agents engage in and the spread of the disease itself. The research shows that it was possible to produce a prototype of a geographically explicit simulation for the Enschede region that is able to simulate 60,000 agents, at a time interval of 30 minutes for a longer period of time (month). The current spread of the disease shows a reproduction rate that is comparable to values found in literature. The geographic spread of the model output (diffusion of the disease) is good.

The social interaction model as described in the conceptual design is based on the concepts of “location”, “time” and “group” where there is full interaction between individuals that are at the same location at the same time but also participate in the same group. Compared to ISTAM (Yang, Atkinson *et al.* 2008) this model is simpler as it does not implement the role of the person participating in the group to the same level of detail. The choice for full participating within the group was because children have very direct contact with each other. They hug classmates and are very intimate with their parents, and do not play a very specific role other than being a child. The selected model is probably over simplified for the adult agents.

The simulation that is presented in this research is geographically explicit in that it assigns a specific location to activities and distributes the population according to reality. It is not geographically specific in simulating the movement between the activity locations. For children in the Netherlands, simulating movement to schools in itself is not relevant. Children in elementary school either walk to school or use a bicycle and it is unlikely that the disease is transmitted during this movement. For the parents (adults) that go to work, when they use private means of transportation (car, bicycle), the risk of disease transmission is also negligible. Adults using public transportation do run a risk of infection during movement.

Yet, within given results of this model, it is still important to simulate this disease in a geographically explicit way because of individuals are spreading the disease while they are moving from place to another. In other word, the mobility of individual and their needs (economic, social etc) to participate in different activities determine which places are in higher risk than others as it can show the geographical trend patterns and the hot spots of disease diffusion.

The City of Enschede was selected as a study area. This city is a suitable study area for a pertussis model as it is large enough to meet the CCS (critical community size). The CCS is the minimal population size below which a disease is unable to maintain itself without external input (Broutin, Simondon *et al.* 2004). Studies indicate the CCS value to be around 250,000 population. The City of Enschede and Hengelo together are in this order of magnitude.

Pertussis is a disease knows to show hierarchical spread, meaning that the disease is likely to spread from urban areas to the surrounding rural areas. Underlying mechanism of this spread is the fact that people commute for work and schooling in the large central city, take the disease to their hometown in the surrounding areas. However, the total population of the villages is below the CCS and the disease dies out in the rural areas. Currently the map extent of the simulation only includes part of the Enschede area. No workers come into the city, and although some agents work outside Enschede, they cannot get infected at their workplaces as they are not modelled explicitly. The school system is currently also a closed system. It would be very interesting enlarge the map extent to include the complete Twente region and see if hierarchical spread can be simulated.

Data used for this simulation posed some limitation on the model implementation. These limitations were found in census data, spatial data and the disease data. For the census data, the limitations relate to the creation of synthetic population. For instance, the data that has been provided for the different age groups only provided information per age group of range 5 years. No exact information about each age size was available. This caused the problem of having the size of a particular age greater than the other ages within the group and that reflected on the produced age groups of children in the simulation. Moreover, employment data of the city population will be useful for the simulation. They will provide the model with the more precise information about the percentage of the workers within the city and the percentage of the worker who are working outside the city. They might be a source of the epidemic in the city or the diffuse the disease from the city to their working destination.

The spatial data did not contain information about location used as activity places. It required addition work to find some crucial locations such as schools or social places. That led the model to assume some places as social places without relating that to the reality.

The empirical disease data did not provide any information about the age profile of recorded cases, the severity of illness, and the value of reproduction ratio that caused that amount of infections. Even with the availability of information about the zip-code of recorded data but it did not provide information whether the zip-code is of the place where the disease occurs or it is the place where the case is registered. It should be stated that the fact that only about 15% of new cases and 3% of re-occurrence cases are registered. This means that in reality the number of pertussis cases is about 9 times as large as what is registered,

Within the simulation itself, couple of limitation existed due to the limitation of the research. The synthetic population is constant no new births were added to the simulation as well as no persons is died during the simulation. The model agents (population) were limited in only families with children. In addition, only the active category of the society was simulated. The older individual in the model is 55 years old and s/he is a parent. That means limited interaction an individual will have. Moreover, the excluded people for the simulation play a role in the disease spread too such as teacher in the schools, health worker etc. who have influence on spreading the disease.

Diversity of agents is too small. It is likely that the behaviour of agents differ based on different cultural backgrounds, unemployment, introducing elderly agents may bring new behaviour to the simulation. Because of limitation of time and data this has been ignored. The population had limited diversity. It existed only among the different age groups but one the level of the macro data population has common behaviours. However, adding diversity of agents would be a good further step for the model.

The city environment is constant. People did not move house or change their activity places such as changing the school, workplace etc. Moreover, the movement of agents from location to another is virtual movement. They did not use the road for their moving and changing their activity locations.

Activity patterns vary during the week for an individual but the weekly pattern of an individual is static over different next weeks. The first week is copied to next weeks. The weekend pattern of the entire city population is the same. All the families follow the same pattern of activity but might be in different locations. In addition, there is no consideration about the vacations during the simulation time, as it means no schools people usually travel to some other places. It has shown in the empirical data that most cases where in school time. However there are some picks in summer too. That might because of the other social activity that the children attend rather than schools also can effect on the results at the end.

The disease model used is also relatively simple. It does not contain any stochastic or mathematical method determining the probability of disease transmission. It was limited on engaging individuals within groups and the possibility if the susceptible and infected individuals existed in the same group. The contact within the group or between groups was not simulated. Furthermore, the illness model was not simulated. Simulating this illness timeline is important as it has an effect on the parent activity who are going to stay at home taking care of their children and again back to the work after recovery. In addition, the recovered individual will considered has high immunity level after infecting and recovering.

There is no reaction when the disease occurs except from staying at home. The individual after entering the period of illness, stayed at home with no activities until the end of the simulation. Further reactions are not simulated such as isolation the infected individual when s/he becomes contagious or even after become ill there is no medical care. Moreover, the illness until recovery stages was not implemented.

During the epidemic, no preventive measures were implemented. Finally, no controller agent exists such as the “government” agent who will “wakes up” when the outbreak start preventing people to engage with group activities where the risk is predicted in some places rather than others.

For better analyzing the spread of the disease, social networks are needed to be visualized using the ABM toolkit RepastS. These social networks if there were there they could help in tracking the disease diffusion from “cloud” group of people to another. The model provided a temporal solution for this absence which was the use of GIS spider tool and the Time-Cube tool to visualizing the data of tracking the disease pattern. Besides there nice way of visualizing the results, they caused some small problems related to their way of dealing with data.

7.2. Recommendations

The model built the synthetic population of the city, provided agents with activity patters; let them engaged in groups and the result of that pertussis disease spread spatially from initially one or two individuals to many of them. However, the work still needs further steps of developing it.

The synthetic population needs to increase the diversity of its components by including different individual categories such as elder people, different ethnic background, and people come from outside the city and working. These aspects of population may add different patterns to the disease spread.
The spatial data need more data details about the capacity of building and the number of departments with a building. That will help in simulating and distributing individuals and their activity groups within the location more accurately.

Simulate the individual activities with more diversity. Within the proposed model, an individual has only 4 to 5 activity per day and during night there are no activities. That off course not as realistic as people used to have. Usually people have different activities during day and night. Disease spread during night as well as during the day. Here the focus will be one the activities of adults more than children. Moreover, for the current time being the wake up time for all the individuals start at the same moment, this is not quite true for the reality. Going for work, schools, travelling to other cities are varying among people and on the city level there is no unique time to start the day.

Spreading diseases is a complex process not easy to be simulated perfectly. Things when related to biological staff it is not easy to deal with simply. The diffusion of a disease depends on many aspects that might not be found within this model or even other. Modeller took some aspects of the whole process and trying to provide them with good explanation or tools helps for more explanation. However, they used to get help, hint and continuing the related previous works just to add more accuracy, efficiency to the subject of simulating. That is what happened within this model. We built the model based on the principles provided by ISTAM model and vaccination model but still in the aspect of "pertussis spread" need further work. Aspects such as the rate of contact for an individual per day, frequency of meeting, simulating the "social distance" between contacted person in addition to the biology behaviour of both individuals and disease are strongly recommended to be included within this model. It was the matter of understanding the pressure of time limitation otherwise; it would be very nice complete project if these aspects were in consideration. However, in scientific researches, everyone tries to put a block on the wall of science for serving humanity and the sustainability of life.
Beyond the limitation of time, the model required to be provided with social networks of its individuals. There is a framework called ReSoNaTa presented by (Holzhauer 2009), which is used to visualize and measure the social networks within Repast platform.

Finally, it would be good if the model be used with another case study whether by choosing another city to apply the model on or apply the model simulate the spread of another disease. The initial and essential aim of designing the model is to build a generic model applicable for other infectious diseases.

Reference

Águas, Gonçalves, *et al.* (2006). "Pertussis: Increasing Disease as a Consequence of Reducing Transmission." The Lancet Infectious Diseases 6(2): 112-117.

Anderson and May (1979). "Population Biology of Infectious Diseases: Part I." Nature 280(5721): 361-367.

Aquino (2003). "Jts Topology Suite Developer's Guide." VIVID Solutions.

Barrett, Eubakn, *et al.* (2005). "If Smallpox Strikes Portland." Scientific American.

Beckman, Baggerly, *et al.* (1996). "Creating Synthetic Baseline Populations." Transportation Research Part A: Policy and Practice 30(6): 415-429.

Bentsi-Enchill, Scott A. Halperin, *et al.* (1997). "Estimates of the Effectiveness of a Whole-Cell Pertussis Vaccine from an Outbreak in an Immunized Population " Vaccine 15(3): 301-306.

Bian (2004). "A Conceptual Framework for an Individual-Based Spatially Explicit Epidemiological Model." Environment and Planning B: Planning and Design 31(3): 381-395.

Bonabeau (2002). "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems." Proceedings of the National Academy of Sciences of the United States of America 99(Suppl 3): 7280-7287.

Bonten, Austin, *et al.* (2001). "Understanding the Spread of Antibiotic Resistant Pathogens in Hospitals: Mathematical Models as Tools for Control." Clinical Infectious Diseases 33(10): 1739-1746.

Boven, Hester E. de Melker, *et al.* (2000). "Waning Immunity and Sub-Clinical Infection in an Epidemic Model: Implications for Pertussis in the Netherlands." Mathematical Biosciences 164(2): 161-182.

Broutin, Simondon, *et al.* (2004). "Whooping Cough Metapopulation Dynamics in Tropical Conditions: Disease Persistence and Impact of Vaccination." Proceedings of the Royal Society of London. Series B: Biological Sciences 271(Suppl 5): S302-S305.

Brown, Riolo, *et al.* (2004). "Spatial Process and Data Models: Toward Integration of Agent-Based Models and Gis." Journal of Geographical Systems 7((1)): 25-47. .

Carley (2009). "Computational Modeling for Reasoning About the Social Behavior of Humans." Computational & Mathematical Organization Theory 15(1): 47-59.

Carley, Altman, *et al.* (2004). "Biowar: A City-Scale Multi-Agent Network Model of Weaponized Biological Attacks Casos Technical Report." CMU-ISRI-04-101.

Castle and Crooks (2006). "Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations " UCL Centre for Advanced Spatial Analysis

Cherry (1999). "Epidemiological, Clinical, and Laboratory Aspects of Pertussis in Adults." Clinical Infectious Diseases 28(s2): S112-S117.

Christie, Marx, *et al.* (1994). "The 1993 Epidemic of Pertussis in Cincinnati -- Resurgence of Disease in a Highly Immunized Population of Children." N Engl J Med 331(1): 16-21.

Claude , Perrin, *et al.* (2009). Considerations for a Social and Geographical Framework for Agent-Based Epidemics. Proceedings of the 2009 International Conference on Computational Aspects of Social Networks, IEEE Computer Society.

Claude, Perrin, *et al.* (2009). Considerations for a Social and Geographical Framework for Agent-Based Epidemics. Proceedings of the 2009 International Conference on Computational Aspects of Social Networks, IEEE Computer Society.

Cointet and Roth (2007). "How Realistic Should Knowledge Diffusion Models Be?" Journal of Artificial Societies and Social Simulations JASSS, 10(3).

Colizza, Barthélemy, *et al.* (2007). "Epidemic Modeling in Complex Realities." Comptes Rendus Biologies 330(4): 364-374.

Collier (2002). "Repast: An Extensible Framework for Agent Simulation." Social Science Research Computing, University of Chicago.

Connell, Dawson, *et al.* (2009). "Comparison of an Agent-Based Model of Disease Propagation with the Generalised Sir Epidemic Model." Air Operations Division Defence Science and Technology Organisation.

Crooks (2007). "The Repast Simulation/Modelling System for Geospatial Simulation." UCL Centre for Advanced Spatial Analysis.

Crowcroft and Pebody (2006). "Recent Developments in Pertussis." Lancet 367: 1926-36personate.

Cullen and Godson (1975). "Urban Networks: The Structure of Activity Patterns." Progress in Planning 4(Part 1): 1-96.

Dalton, Durrheim, *et al.* (2008). "Likely Impact of School and Childcare Closures on Public Health Workforce During an Influenza Pandemic: A Survey." CDI 32(2).

De Melker, Spaendonck, *et al.* (1997). "Pertussis in the Netherlands: An Outbreak Despite High Levels of Immunization with Whole-Cell Vaccine." Emerging Infectious Diseases 3(2);.

de Vries "Modelling Malaria Risk: An Individual Based and Spatial Explicit Approach." workshop Spatial Aspects of Demography.

Eames and Keeling (2002). "Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases." Proceedings of the National Academy of Sciences of the United States of America 99(20): 13330-13335.

Eames, Read, *et al.* (2009). "Epidemic Prediction and Control in Weighted Networks." Epidemics 1(1): 70-76.

Edwards and Halasa (2004). "Commentary: Is Pertussis Disease Increasing?" International Journal of Epidemiology 33: 365-366

Ellegård (1999a). "A Time-Geographical Approach to the Study of Everyday Life of Individuals- a Challange of Complixity." GeoJournal 48: 167 - 175.

Ellegård (1999b). "A Time-Geographical Approach to the Study of Everyday Life of Individuals – a Challenge of Complexity." GeoJournal 48(3): 167-175.

Enschede-Municipality. (2009). "Enschede in Figures: Statistical Yearbook." from http://www.enschede.nl/gemeente/cijfers/statjaarboek/.

Epstein (1999). "Agent-Based Computational Models and Generative Social Science " Economic Studies, Brookings Institution, Washington 4(5).

Epstein and Axtell (1996). Growing Artificial Societies: Social Science from the Bottom Up, The MIT Press.

Epstein, Cummings, *et al.* (2004). "Toward a Containtment Strategy for Smallpox Bioterror: An Individual-Based Computational Approach." Brookings Institution Press, Washington D.C.

Ettema, Arentze, *et al.* (2009). "Modelling the Influence of Temporal and Monetary Constraints on Activity Participtaion, Travel, Consumption of Goods, Residential Location and Work Status: Application in Land-Use Transport Interaction Model."

Ettema, Jong Kor de, *et al.* (2005) "Puma: Multi-Agent Modelling of Urban Systems."

Ettema and Timmermans (1997). Activity-Based Approaches to Travel Analysis.

Eubakn (2005). "Network Based Models of Infectious Disease Spread." Japanese journal of infectious diseases (Jpn J Infect Dis)

vol 58 (6): S9-13.

Eubank, Guclu, *et al.* (2004). "Modelling Disease Outbreaks in Realistic Urban Social Networks." Nature 429(6988): 180-184.

Ferguson, Cummings, *et al.* (2006). "Strategies for Mitigating an Influenza Pandemic." Nature 442(7101): 448-452.

Forsyth, Nagai, *et al.* (2005). "Pertussis Immunization in the Global Pertussis Initiative International Region: Recommended Strategies and Implementation Considerations." The Pediatric Infectious Disease Journal 24(5): S93-S97.

Galvani and May (2005). "Epidemiology: Dimensions of Superspreading." Nature 438(7066): 293-295.

Gilberg, Elisabeth Njamkepo, *et al.* (2002). "Evidence of Bordetella Pertussis Infection in Adults Presenting with Persistent Cough in a French Area with Very High Whole-Cell Vaccine Coverage." The Journal of Infectious Diseases 186(03): 415-418.

Gilbert and Terna (2000). "How to Build and Use Agent-Based Models in Social Science." Mind & Society 1: 57-72.

Gilbert and Troitzsch (2005). Simulation for the Social Scientist, Open University Press.

Grenfell, Bjornstad, *et al.* (2002). "Dynamics of Measles Epidemics: Scaling Noise, Determinism, and Predictability with the Tsir Model." Ecological Monographs 72(2): 185-202.

Grimm, Berger, *et al.* (2006). "A Standard Protocol for Describing Individual-Based and Agent-Based Models." Ecological Modelling 198(1-2): 115-126.

Gu and Novak (2009). "Agent-Based Modelling of Mosquito Foraging Behaviour for Malaria Control." Transactions of the Royal Society of Tropical Medicine and Hygiene 103(11): 1105-1112.

Güriş, Peter M. Strebel, *et al.* (1999). "Changing Epidemiology of Pertussis in the United States: Increasing Reported Incidence among Adolescents and Adults, 1990-1996." Clinical Infectious Diseases 28(6): 1230-1237.

H. E. de Melker, M. A. Conyn-van Spaendonck, *et al.* (1997). "Pertussis in the Netherlands: An Outbreak Despite High Levels of Immunization with Whole-Cell Vaccine." Emerging Infectious Diseases 3(2);.

HCN (2004) "Health Council of the Netherlands. Vaccination against Pertussis. ." publication no. 2004/04e.

Hester E. de Melker, J.F.P. Schellekens, *et al.* (2000). "Reemergence of Pertussis in the Highly Vaccinated Population of the Netherlands: Observations on Surveillance Data." Emerging Infectious Diseases 6.

Hethcote (1997). "An Age-Structured Model for Pertussis Transmission." Mathematical Biosciences 145(2): 89-136.

Hethcote (1999a). "Simulations of Pertussis Epidemiology in the United States: Effects of Adult Booster Vaccinations." Mathematical Biosciences 158: 47-73.

Hethcote (1999b). "Simulations of Pertussis Epidemiology in the United States: Effects of Adult Booster Vaccinations." Mathematical Biosciences 158(1): 47-73.

Hethcote (2000). "The Mathematics of Infectious Diseases." Society for Industrial and Applied Mathematics 42(4): 599 - 653.

Hethcote and Yorke (1984). Gonorrhea Transmission Dynamics and Control. Berlin; New York, Springer-Verlag.

Hoey (2003). "Pertussis in Adults." Canadian Medical Association Journal 168(4).

Holzhauer (2009). "Developing a Socail Network Analysis and Visualization Module for Repast Models." CESR.

Howe, Collier, *et al.* (2006). Containing Agents: Contexts, Projections and Agents. Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects. Sallach, Macal and North. Argonne National Laboratory and The University of Chicago.

http://repast.sourceforge.net/docs/reference/SIM/ "Repast S Documents." Repast Simphony : Working with the scheduler.

Huang and Williamson (2001/2). "A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-Area Microdata." working paper. Department of Geography, University of Liverpool.

Huhns and Singh, Eds. (1998). Readings in Agents, Morgan Kaufmann Publishers Inc.

J. Ryan, H. Maoh, *et al.* (2007). ""Population Synthesis:Comparing the Major Techniques Using a Small, Complete Population of Firms"." Centre for Spatial Analysis,McMaster University,Hamilton, Canada.(Geographical Analysis).

JavaSun. (2010). "Object-Oriented Programming Concepts." from http://java.sun.com/docs/books/tutorial/java/concepts/object.html.

Jennings, Sycara, *et al.* (1998). "A Roadmap of Agent Research and Development." Autonomous Agents and Multi-Agent Systems 1(1): 7-38.

Juliet Nakamya, Elke Moons, *et al.* (2009). "How Real Are Synthetic Population?" Transportation Research Institute, Hasselt University, Belgium.

Keeling and Eames (2005). "Networks and Epidemic Models." Journal of the Royal Society, Interface / the Royal Society vol. 2 no. 4: 295-307

Kohler and Gumerman (2000). Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes, Oxford University Press.

Lee, Brown, *et al.* (2009). "Simulating School Closure Strategies to Mitigate an Influenza Epidemic." Journal of Public Health Management and Practice Publish Ahead of Print: 10.1097/PHH.0b013e3181ce594e.

Lee, LeBaron, *et al.* (2005). "Pertussis in Adolescents and Adults: Should We Vaccinate?" Pediatrics 115(6): 1675-1684.

Liebert "Creating a Basic Gis Agent Environment."

Macal and North (2005). Tutorial on Agent-Based Modeling and Simulation. Proceedings of the 2005 Winter Simulation Conference. Kuhl, Steiger, Armstrong and Joines. USA.

Malleson (2008). "Repast Simphony “No-Gui” Tutorial."

Markel, Lipman, *et al.* (2007). "Nonpharmaceutical Interventions Implemented by Us Cities During the 1918-1919 Influenza Pandemic." JAMA 298(6): 644-654.

Netherlands (2004) "Vaccination against Pertussis. The Hague: Health Council of the Netherlands." publication no. 2004/04e.

Newman (2002). "Spread of Epidemic Disease on Networks." Physical Review E 66(1): 016128.

North, Howe, *et al.* (2005a). The Repast Simphony Development Environment. Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms Macal, North and Sallach. Argonne National Laboratory and The University of Chicago.

North, Howe, *et al.* (2005b). The Repast Simphony Runtime System. Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms,. Macal, North and Sallach. Argonne National Laboratory and The University of Chicago, .

North, Howe, *et al.* (2007). A Declarative Model Assembly Infrastructure for Verification and Validation. Advancing Social Simulation: The First World Congress**:** 129-140.

North, Nicholson, *et al.* (2006). "Experiences Creating Three Implementations of the Repast Agent Modeling Toolkit." ACM Transactions on Modeling and Computer Simulation 16,(1): 1-25.

Oppen-Stuyt (2009). "A Brief Outline of the Educational System in the Nl." International Education Solution House - Educaide

Perez and Dragicevic (2009). "An Agent-Based Approach for Modelling Dynamics of Contagious Disease Dpread." International Journal of Health Geographics 8(50).

Railsback, Lytinen, *et al.* (2006). "Agent-Based Simulation Platforms: Review and Development Recommendations." Simulation -California- 82(9): 609-623

Read, Eames, *et al.* (2008). "Dynamic Social Networks and the Implications for the Spread of Infectious Disease." J. R. Soc. Interface **5**: 1001-1007.

Riley, Fraser, *et al.* (2003). "Transmission Dynamics of the Etiological Agent of Sars in Hong Kong: Impact of Public Health Interventions." Science 300(5627): 1961-1966.

Ropella, Railsback, *et al.* (2008). "Software Engineering Considerations for Individual-Based Models." Natural Resource Modeling 15(1): 5 - 22.

Russell and Norvig, Eds. (2003). Artificial Intelligence: A Modern Approach, Prentice Hall.

Sansores and Pavón (2005). "Agent Based Simulation for Social Systems: From Modeling to Implementation." Universidad Complutense de Madrid, Dep. Sistemas Informáticos y Programación.

Sargent (2007). Verification and Validation of Simulation Models Proceedings of the 2007 Winter Simulation Conference

Skowronski, De Serres, *et al.* (2002). "The Changing Age and Seasonal Profile of Pertussis in Canada." The Journal of Infectious Diseases 185(10): 1448-1453.

Taranger, Trollfors, *et al.* (2001). "Immunologic and Epidemiologic Experience of Vaccination with a Monocomponent Pertussis Toxoid Vaccine." Pediatrics 108(6): e115-.

Tatara, North, *et al.* (2006). An Introduction to Repast Simphony Modeling Using Simple Predator-Pery Example. Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects. Sallach, Macal and North. The University of Chicago, Argonne National Laboratory and The University of Chicago.

Teweldemedhin, Marwala, *et al.* (2004). "Agent-Based Modelling: A Case Study in Hiv Epidemic." Proceedings of the Fourth International Conference on Hybrid Intelligent Systems.

Timmermans and Zhang (2009). "Modeling Household Activity Travel Behavior: Examples of State of the Art Modeling Approaches and Research Agenda." Transportation Research Part B: Methodological 43(2): 187-190.

Tobias and Hofmann (2004). "Evaluation of Free Java-Libraries for Social-Scientific Agent Based Simulation." Journal of Artificial Societies and Social Simulation 7(1).

Toroczkai and Guclu (2007). "Proximity Networks and Epidemics." Physica A: Statistical Mechanics and its Applications 378(1): 68-75.

Tuckel, Sassler, *et al.* (2006). "The Diffusion of the Influenza Pandemic of 1918 in Hartford, Connecticut." Social Science History 30(2): 167-196.

Van Boven, De Melker, *et al.* (2000). "Waning Immunity and Sub-Clinical Infection in an Epidemic Model: Implications for Pertussis in the Netherlands." Mathematical Biosciences 164(2): 161-182.

Wendelboe, Njamkepo, *et al.* (2007). "Transmission of Bordetella Pertussis to Young Infants." The Pediatric Infectious Disease Journal 26(4).

Wey, Blumstein, *et al.* (2008). "Social Network Analysis of Animal Behaviour: A Promising Tool for the Study of Sociality." Animal Behaviour 75(2): 333-344.

WHO-Website (2010). Pandemic (H1n1) 2009 - Update 82.

WHO (2007). "Laboratory Manual for the Diagnosis of Whooping Cough Caused by Bordetella Pertussis/Bordetella Parapertussis." Department of Immunization, Vaccines and Biologicals. Geneva, Switzerland.

Wirsing von König CH, S Halperin, *et al.* (2002). "Pertussis of Adults and Infants." The Lancet Infectious Diseases 2(12): 744-750.

Wooldridge and Jennings (1995). "Intelligent Agents: Theory and Practice." Knowledge Engineering Review 10(2).

Yang and Atkinson (2007). "Individual Space -Time Activity-Based Model: A Model for the Simulation of Airborne Infectious-Disease Transmission by Activity-Bundle Simulation." Environment and Planning B: Planning and Design.

Yang and Atkinson (2008). "Individual Space – Time Activity-Based Model: A Model for the Simulation of Airborne Infectious-Disease Transmission by Activity-Bundle Simulation." Environment and Planning B: Planning and Design

35((1)): 80 - 99

Yang, Atkinson, *et al.* (2008). "Individual Space-Time Activity-Based Modelling of Infectious Disease Transmission within a City." J. R. Soc. Interface 5: 759-772.

Yong Yang and Atkinson (2008). ":Individual Space – Time Activity-Based Model: A Model for the Simulation of Airborne Infectious-Disease Transmission by Activity-Bundle Simulation." Environment and Planning B: Planning and Design

35((1)): 80 - 99

S. G. Henderson, B. Biller, M.-H. Hsiehet al.